FALSE NARRATIVES AND POLITICAL MOBILIZATION

Kfir Eliaz
Tel Aviv University, Israel, University of Utah, USA

Simone Galperti
UC San Diego, USA

Ran Spiegler
Tel Aviv University, Israel, University College London, UK

Abstract
We present an equilibrium model of politics in which political platforms compete over public opinion. A platform consists of a policy, a coalition of social groups with diverse intrinsic attitudes to policies, and a narrative. We conceptualize narratives as subjective models that attribute a commonly valued outcome to (potentially spurious) postulated causes. When quantified against empirical observations, these models generate a shared belief among coalition members over the outcome as a function of its postulated causes. The intensity of this belief and the members' intrinsic attitudes to the platform's policy determine the extent to which the coalition is mobilized. Only platforms that generate maximal mobilization prevail in equilibrium. Our equilibrium characterization demonstrates how false narratives can be detrimental to the commonly valued outcome, and how political fragmentation leads to their proliferation. The false narratives that emerge in equilibrium have a flavor of "scapegoating:" They attribute good outcomes to the exclusion of social groups from ruling coalitions. (JEL: D72, D74, D83, P00)

1. Introduction

Success in democratic politics requires the mobilization of public opinion, which takes various forms: rallies, petitions, social media activism, and ultimately voter turnout. Shifts in public opinion can explain which policies get implemented and which

The editor in charge of this paper was Nicola Pavoni.

Acknowledgments: Spiegler acknowledges financial support from ERC Advanced Investigator grant no. 692995 and Leverhulme Trust grant no. RPG-2023-120. We thank Gianpaolo Bonomi, Tuval Danenberg, Danil Dmitriev, Nathan Hancart, Federica Izzo, Gilat Levy, Guido Tabellini, numerous seminar participants, and the editor and referees of this journal, for helpful comments.

E-mail: kfire@tauex.tau.ac.il (Eliaz); sgalperti@ucsd.edu (Galperti); rani@post.tau.ac.il (Spiegler)

Journal of the European Economic Association 2025 23(3):983–1027
https://doi.org/10.1093/jeea/jvae047

coalitions of social groups form around them (Burstein 2003). In turn, opinion makers (politicians, news outlets, and pundits) use past performance of policies and coalitions as raw material for shaping public opinion. This paper is an attempt to shed light on this interplay.

Our starting point is the idea that *narratives* are a powerful tool for mobilizing public opinion. This is a familiar idea with numerous expressions in academic and popular discourse. After Senator John Kerry lost the 2004 presidential elections, his political strategist Stanley Greenberg said that "a narrative is the key to everything" and that Republicans had "a narrative that motivated their voters".[1] Shanahan, McBeth, and Hathaway (2011) write: "Policy narratives are the lifeblood of politics. These strategically constructed 'stories' contain predictable elements and strategies whose aim is to influence public opinion toward support for a particular policy preference". And Stone (1989) writes:

> "... political actors use narrative story lines... to manipulate so-called issue characteristics... As one side in a political battle seeks to push a problem into the realm of human purpose, the other side seeks to push it away from intent toward the realm of nature or to show that the problem was intentionally caused by someone else."

This paper is a theoretical study of how narratives shape public-opinion battles in heterogeneous societies. We explore what makes narratives more or less popular, and what role they play in the determination of policies and the formation of ruling coalitions.

We formalize political narratives as *causal models* that attribute public outcomes (e.g., economic growth) to postulated causes. Echoing the quote from Stone (1989), these causes can be policies (e.g., attributing growth to economic policy), governing parties (e.g., attributing growth to whether Democrats or Republicans were in power—without getting into the specific policies they implemented while in power), or external elements beyond governments' control (e.g., attributing growth to technological shocks). By this view, a false narrative is a misspecified causal model that attributes outcomes to wrong causes.

In our model, a narrative generates a probabilistic belief regarding the effect of a postulated cause on the outcome by "estimating" the empirical correlation between them. A false narrative can produce wrong beliefs by assigning an incorrect causal meaning to the correlation it highlights. The stronger this correlation, the stronger the causal belief that the narrative generates—which translates into greater mobilization of social groups behind the political platform employing that narrative. Thus, competition between platforms for public support is, to some extent, a battle between conflicting narratives over what drives public outcomes.

We consider a heterogenous society that consists of multiple social groups having different intrinsic attitudes to policies. We think of a social group as a collection of agents with shared ideological, socioeconomic, or ethnic/religious characteristics, as well as a *distinct* political representation (in line with Lipsett and Rokkan's (1967)

1. See William Safire's New York Times article titled "Narrative" (https://www.nytimes.com/2004/12/05/magazine/narrative.html).

"cleavage theory" according to which, there is a fixed mapping between voting blocs and political parties). For example, society can be divided into left and right wings, possibly with finer subdivisions. Other examples include the Flemish and French parties in Belgium, or the various ethnic and religious parties in Israel.

We make the simplifying assumption that policies are the *only true cause* of public outcomes. The differences between the intrinsic policy attitudes of social groups will naturally give rise to correlations between the structure of ruling coalitions, the policies they implement, and these policies' outcomes. A false narrative can exploit these correlations and causally attribute the outcome *solely* to a social group's power status (i.e., whether it belongs to the ruling coalition), even though in reality this correlation is due to confounding by the implemented policies.

For illustration, suppose coalition C usually refrains from taxing wealth. As a result, social inequality tends to rise when C is in power. A rival coalition C' may exploit this correlation and spin a false narrative that, in order to reduce inequality, we only need to keep the social groups behind C out of power. Because this narrative does not attribute the outcome to its true cause (namely, tax policy), it enables C' to gain support. On one hand, C' can act exactly like C by not proposing an unpopular wealth tax. On the other hand, it can claim that by elbowing out C it is doing something to lower inequality, which is popular. Thus, in a sense, C' uses C as a "scapegoat" to hide the link between an attractive policy and its unattractive consequences. Our main objective in this paper is to understand how such false narratives can gain ascendancy, what form they take, and how they shape public policies and ruling coalitions.

In our setting, a policy, a coalition of social groups, and a narrative form a *political platform*. Given a long-run joint empirical distribution over prevailing platforms and public outcomes, different narratives may induce conflicting beliefs regarding the consequences of policies and coalitions. The long-run frequencies of prevailing platforms and outcomes affect narrative-based causal beliefs, which (through their effect on political mobilization) determine the platforms that prevail. This feedback effect suggests a need for an *equilibrium* notion of prevailing political platforms.

We define an equilibrium as a probability distribution over prevailing platforms, such that every platform in its support maximizes the total mobilization of the social groups belonging to the platform's coalition. This definition captures the idea that a platform's success depends on the strength of its popular support (in terms of the number and size of participating social groups as well as the intensity of their participation). It does so in the spirit of competitive equilibrium, as in Rothschild and Stiglitz (1976). The backstory is that there is "free entry" of office-motivated political entrepreneurs who propose policy-narrative combinations. If a particular combination attracts stronger support than the current combination, the former will overthrow the latter. Eventually, the platform that maximizes total support will prevail.[2] One advantage of our approach is that it avoids the nitty–gritty of modeling the formation

2. Section 3 illustrates such a dynamic process and Section 6.2 leverages it to offer a foundation to our equilibrium concept.

of parliamentary coalitions (which is only partly related to battles over public opinion, our main concern here).

Using this formalism, we obtain several insights. First, in addition to the true narrative that attributes outcomes to policies, two types of false narratives emerge in equilibrium, in a way that echoes the above quote from Stone (1989). The first type is a "denial" narrative that does not attribute outcomes to any endogenous variable (thus implicitly attributing it to external forces). The other type is a "tribal" narrative that attributes a good public outcome to the *exclusion* of some social groups from the ruling coalition. In a political speech or a social-media post, such a narrative could appear as "national security is strong when the Left is out of power."

Recent public debates over high inflation, which have involved competing claims over its causes, are suggestive of these types of narratives. Some narratives attribute inflation to government actions (fiscal expansion), others to external factors (global supply-chain disruptions), and yet others assign credit or blame solely to the party in power, without attempting to link inflation to the party's policies. A selection of press quotes demonstrates the form of these conflicting narratives:

> "As prices have increased ... some Democrats have landed on a new culprit: price gouging ... For Democrats, it is a convenient explanation as inflation turns voters against President Biden. It lets Democrats deflect blame from their pandemic relief bill, the American Rescue Plan, which experts say helped increase prices."[3]
>
> "Democrats have blamed supply chain deficiencies due to COVID-19, as well as large corporations and monopolies."[4]
>
> "As the midterm elections draw nearer, a central conservative narrative is coming into sharp focus: President Biden and the Democratic-controlled Congress have made a mess of the American economy." [5]

The distinction between a false narrative that attributes outcomes to whoever is in power and a more accurate narrative that attributes outcomes to policies appears in Paul Krugman's recent article about the politics of inflation:

> "... voters aren't saying, 'Trimmed mean P.C.E. inflation is too high because fiscal policy was too expansionary'. They're saying, 'Gas and food were cheap, and now they're expensive ..'. So when people say—as they do—that gas and food were cheaper when Donald Trump was president, what do they imagine he could or would be doing to keep them low if he were still in office?"[6]

Our second insight is that the false narratives employed in equilibrium sustain policies that would not be taken if the only prevailing narrative were the true one (which correctly attributes outcomes to policies). The function of false narratives is to resolve the cognitive dissonance between the intrinsic appeal of a policy and its objective

3. https://www.nytimes.com/2022/06/14/briefing/inflation-supply-chain-greedflation.html.

4. https://fivethirtyeight.com/features/what-democrats-and-republicans-get-wrong-about-inflation/.

5. https://www.nytimes.com/2022/06/11/opinion/fed-federal-reserve-inflation-democrats.html.

6. https://www.nytimes.com/2022/06/02/opinion/inflation-biden.html. See also Weaver (2013) and Sanders, Hurtado, and Zoragastua (2017).

inadequacy for the desired outcome. This is achieved by deflecting responsibility for the outcome from its true cause to spurious causes.

Moreover, when society becomes more politically fragmented (in the sense that finer social groups have distinct political representation), tribal narratives proliferate and can lead to further crowding out of the true narrative and the policy it justifies. Greater polarization of attitudes toward policies has a similar equilibrium effect. We illustrate these points in a setting where social groups and tribal narratives are defined by a collection of binary attributes.

Finally, we characterize the structure of coalitions that form in equilibrium. False narratives give rise to coalitions that would not form if only the true narrative prevailed. In particular, when a political platform employs a tribal narrative, it excludes social groups that do not oppose the platform's policy (indeed, they implement the same policy when they are in power). While this exclusion shrinks the coalition and might therefore seem to hurt its mobilization, it has the compensating effect of strengthening the causal belief that the tribal narrative generates. Thus, our results suggest that the mobilizing power of false tribal narratives has substantial implications for implemented policies and prevailing social coalitions.

2. A Model

We begin by describing the model's primitives. Let $y \in \{B, G\}$ be a *public outcome*. There is a social consensus that $y = G$ is a "good" outcome. Let $a \in A = \{b, g\}$ be a *policy*. Policies cause outcomes according to the objective conditional probability distribution

$$\Pr(y = G \mid a) = \begin{cases} \dots & \\ q & \text{if } a = g \\ 0 & \text{if } a = b \end{cases}, \tag{1}$$

where $q \in (0, 1]$.

Let $N = \{1, \dots, n\}$ be a set of *social groups*, where $n \geq 2$. A *coalition* is a nonempty subset C of N. Define a function $f : N \times A \to \mathbb{R}_+$. We refer to $f(i, a)$ as group i's *mobilization propensity* given policy a. This reflects group i's intrinsic attitudes toward a. For example, when $y = G$ represents low inflation and g (b) represents fiscal restraint (expansion), $f(i, b) > f(i, g)$ means that group i finds fiscal expansion intrinsically more attractive than fiscal restraint. For all i, $f(i, a) > 0$ for at least one a.

Using these primitives, we now present the key definitions of the model.

Narratives. To formulate our notion of narratives, we introduce a language that encodes policies and coalitions. Let $x = (x_0, \dots, x_n)$ be a profile of binary variables, where $x_0 \in \{b, g\}$ and $x_i \in \{0, 1\}$ for every $i > 0$. Define the following function that assigns values of x to every policy-coalition pair (a, C): $x_0(a, C) = a$, and for $i > 0$, $x_i(a, C) = 1$ if and only if $i \in C$. For instance, if $N = \{1, 2, 3\}$ and $(a, C) = (g, \{2, 3\})$, then $x = (g, 0, 1, 1)$. If C is interpreted as a ruling coalition, the

variable $x_i(a, C)$ encodes the "*power status*" of group i—that is, whether it is part of the ruling coalition. For example, when $(a, C) = (g, \{2, 3\})$, then $x_1(a, C) = 0$ and $x_2(a, C) = 1$.

A *narrative* is a set $S \subseteq \{0, 1, \ldots, n\}$, namely a subset of the components of x. The set S defines the variables to which the outcome y is attributed. For example, $S = \{0, 2\}$ means that the postulated causes of y are the policy and group 2's power status. Given a probability distribution p over (x, y), a narrative S generates a belief over the outcome conditional on its postulated causes. We denote this belief by $(p(y \mid x_S))$, where $x_S = (x_i)_{i \in S}$.[7] Thus, a narrative S draws attention to the correlation between y and x_S and gives this correlation a causal meaning.

We refer to $S = \{0\}$ as the "*true*" narrative, because it attributes y to its sole true cause a. Every narrative that fails to include 0 is false because it attributes y to wrong causes. We refer to $S = \varnothing$ as a "*denial*" narrative because it does not attribute y to any of the endogenous variables. Implicitly, the denial narrative attributes the outcome to external factors. Finally, we refer to non-empty narratives $S \subseteq N$ as "*tribal*" because they attribute y to the power status of social groups, without mentioning policies.

We assume that there is some domain of feasible narratives, which includes the true and denial narratives. We will later consider various domain restrictions.

Platforms and Mobilization. A *platform* is a policy-coalition-narrative triple (a, C, S) with the restriction (to be explained below) that, if $i \in C$, then $f(i, a) > 0$. Let σ denote an objective long-run probability distribution over prevailing platforms (we will clarify below what it means for a platform to prevail.) The induced joint distribution over (a, C, S, y) is

$$p_\sigma(a, C, S, y) = \sigma(a, C, S) \cdot \Pr(y \mid a),$$

where $\Pr(y \mid a)$ is given by (1). We denote the support of σ by $Supp(\sigma)$.

When applied to the distribution $p_\sigma(a, C, S, y)$, a narrative S induces the following conditional belief over y given x:

$$p_\sigma(y \mid x_S) = \sum_a p_\sigma(a \mid x_S) \Pr(y \mid a), \tag{2}$$

where $p_\sigma(a \mid x_S)$ is determined by σ as follows. When $0 \in S$, $p_\sigma(a = x_0 \mid x_S) = 1$. When $0 \notin S$,

$$p_\sigma(a \mid x_S) = \frac{\sum_{C', S' \mid x_S(a, C') = x_S} \sigma(a, C', S')}{\sum_{a', C', S' \mid x_S(a', C') = x_S} \sigma(a', C', S')}.$$

This is the probability that σ assigns to a, conditional on the power status of the groups in S as described by x_S.

7. We use the abbreviated notation $(p(y \mid x_S))$ for $(p(y \mid x_S))_{x_S, y}$.

We assume that the extent to which a platform mobilizes a group is proportional to the promise of a good outcome it offers, where the proportionality constant is the group's mobilization propensity.

DEFINITION 1 (*Mobilization*). Fix a distribution σ over platforms. The extent to which platform (a, C, S) mobilizes group i is

$$m_{i,\sigma}(a, C, S) = p_\sigma(y = G \mid x_S(a, C)) \cdot f(i, a). \tag{3}$$

The term $p_\sigma(y = G \mid x_S(a, C))$ represents a narrative-based probability of a good outcome conditional on the platform—specifically those aspects of the platform that its narrative highlights as relevant causes. It is the empirical frequency of a good outcome (according to the long-run distribution p_σ) conditional on $x_S = x_S(a, C)$.

Equilibrium. We are now ready to define equilibrium in our model. This definition pours content into the notion of prevailing platforms.

DEFINITION 2 (*Equilibrium*). A distribution σ over platforms with full support over (a, C) is an ε-equilibrium if whenever $\sigma(a, C, S) > \varepsilon$, platform (a, C, S) maximizes the total mobilization

$$M_\sigma(a, C, S) = \sum_{i \in C} m_{i,\sigma}(a, C, S). \tag{4}$$

A distribution σ (not necessarily with full support) is an equilibrium if it is the limit of ε-equilibria as $\varepsilon \to 0$.

We start from the notion of ε-equilibrium to ensure that $p_\sigma(y = G \mid x_S)$ is well-defined. This "trembling hand" aspect plays a very limited role in our analysis.

2.1. Discussion

We conclude this section with a discussion of various elements of our model.

The Mobilization Propensity. The function $f(i, a)$ represents in reduced form several aspects of group i: a value judgment of policy a, the policy's specific costs or benefits for the group (independently of its implications for the public outcome), the group's political participation costs and its size. In particular, we can think of an individual social group i as consisting of a mass of agents with distinct attitudes to policies; $f(i, a)$ is the mass of agents in group i who can be mobilized in support of a.

We view $f(i, a) > 0$ and $f(i, a) = 0$ as being qualitatively distinct. This is the reason why our definition of platforms requires that $f(i, a) > 0$ if $i \in C$. Suppose group i is intrinsically *opposed* to policy a. Then, it is natural to assume that this group will not be part of a coalition that advocates a: Either the coalition's gatekeepers will oust what it perceives as a "fifth column", or the group itself would not want to join the coalition in the first place. By assumption, this group satisfies $f(i, a') > 0$ for $a' \neq a$,

so it could join coalitions that advocate a'. In this case, rallying in favor of a' is akin to rallying against a. Modeling political mobilization that consists of protest against a policy without acting in favor of another is outside the scope of this paper.

Group Mobilization. The function M_σ is a measure of the total support that platform (a, C, S) generates, given the distribution σ. Our notion of support takes a broad view of political mobilization to include not only voting, but also other kinds of political participation: rallies, petitions, or social media activism. Expression (4) means that the mobilization of a coalition is proportional to its aggregate mobilization propensity given the platform's policy, as well as to the belief—shaped by the platform's narrative—that the outcome will be good conditional on the event that the platform prevails. The stronger the belief, the stronger the support for the platform.

We adopt the multiplicative form of (3) mainly for tractability. In Section 6.1, we provide " *micro-foundations*" that derive f and m from more elaborate models in which the primitives are *individual preferences*, such that group mobilization arises from members' anticipatory utility from platforms.

We index M_σ by σ because the conditional belief $p_\sigma(y = G \mid x_S)$ may vary with the long-run distribution over prevailing platforms. To see why, recall that y is a fixed (probabilistic) function of only a, so it is independent of C conditional on a. This property can be represented by the directed acyclic graph (DAG) $C \leftarrow a \rightarrow y$.[8] However, if narrative S does not attribute y to a—that is, $0 \notin S$—it amounts to interpreting a long-run correlation between C and y as if it is causal, namely as if the DAG were $x_S \rightarrow y$. In reality, this correlation is due to confounding because both y and C are correlated with a. The latter correlation depends on σ as shown by (2).

We now illustrate how false narratives can induce wrong beliefs about the outcome. Suppose $n = 3$ and σ is as follows:

σ	a	C	S
α	g	$\{1\}$	$\{0\}$
β	b	$\{2,3\}$	$\varnothing$
γ	b	$\{1,3\}$	$\{2\}$.

Then, using (2), we obtain the subjective conditional probability of a good outcome associated with each of the three platforms in $Supp(\sigma)$:

$$p_\sigma(y = G \mid x_{\{0\}}(g, \{1\})) = p_\sigma(y = G \mid a = g) = q$$
$$p_\sigma(y = G \mid x_{\varnothing}(b, \{2,3\})) = p_\sigma(y = G) = q \cdot \alpha,$$

and

$$p_\sigma(y = G \mid x_{\{2\}}(b, \{1,3\}))$$
$$= p_\sigma(y = G \mid x_2 = 0) = p_\sigma(y = G \mid 2 \notin C) = q \cdot \frac{\alpha}{\alpha + \gamma}.$$

8. The link $a \rightarrow y$ represents a true causal relation, whereas the direction of the link between C and a is arbitrary.

For a general distribution σ, the last term would be

$$p_\sigma(y = G \mid x_2 = 0) = \frac{q \sum_{C,S|2\notin C} \sigma(g, C, S)}{\sum_{a,C,S|2\notin C} \sigma(a, C, S)}.$$

Thus, false narratives can generate positive mobilization for platforms that involve policy b, even though it objectively leads to $y = B$ with certainty.

The Equilibrium Concept. Our definition of equilibrium captures the idea that a platform's political power depends on how strongly it mobilizes its coalition groups. We view narrative-fueled political competition as a battle over public opinion. A platform prevails given σ if it generates the largest total mobilization—if it didn't, another platform would arise in the political arena and replace it. When (a, C, S) prevails, C is a *ruling coalition*. The distribution σ describes the long-run frequencies with which different platforms prevail. In Section 6.2, we substantiate this dynamic interpretation of our equilibrium concept.

Note that if only the true narrative $S = \{0\}$ existed, any platform with $a = b$ would generate $M_\sigma = 0$ by (1). Instead, a platform with $a = g$ always generates $M_\sigma > 0$. In this case, policy g would occur with probability one in equilibrium. We therefore refer to g as the "*rational*" policy.

3. Two-Group Societies

We begin our analysis with the simple case of $n = 2$. For concreteness, we present this case in terms of a specific interpretation of our model, which is common in the political economics literature [e.g., Ch. 3 in Persson and Tabellini (2000)]. The outcomes G and B represent successful and failed provision of a public good (more broadly, government functions). The policies g and b represent high and low taxation (more broadly, "big government" vs. "small government"). Our data-generating process means that high taxation is necessary (but insufficient when $q < 1$) for successful public-good provision.[9] The two social groups differ in their attitudes to taxation, because of differences in ideology or income profile. In Section 6.1, we provide a precise formal "*micro-foundation*" for this interpretation.

To avoid trivial cases, we assume $f(1, g) > f(2, g)$ and $f(1, b) < f(2, b)$. That is, group 1's intrinsic support for high (low) taxation is stronger (weaker) than group 2's. We also rule out the grand coalition: C can only be $\{1\}$ or $\{2\}$. This specification is akin to a two-party system, in which exactly one party can be in power at any point in time. In this case, our equilibrium concept can be interpreted in terms of a two-party voting model: Supporters of each party vote non-strategically for it, to the extent that the party's policy-narrative bundle mobilizes them to do so—otherwise, they abstain [somewhat as in Levy, Razin, and Young (2022)].

9. The "leftist" bias of this interpretation will be offset by a "rightist" bias in a later example.

This setting allows us to reduce the set of relevant narratives. Since $x_1 = 1$ if and only if $x_2 = 0$, all tribal narratives $S \subseteq N$ are equivalent. When they accompany the coalition $\{i\}$, they effectively say that the public good tends to be successfully provided when group i is in power (or, equivalently, when group j is not in power). In addition, all S that contain $\{0\}$ are equivalent, because $\Pr(y = G \mid a, C) = \Pr(y = G \mid a)$ for all a, C. Every feasible narrative is then equivalent to one of the following: the true narrative $\{0\}$, the denial narrative $\varnothing$, or the tribal narrative $\{1\}$. Therefore, in this section, we assume that *only these three narratives* are feasible—and we denote them by *true*, *denial*, and *tribal* for expositional clarity. This assumption is without loss of generality as far as the equilibrium distribution over (a, C) is concerned. This de-facto reduction to a two-party model with few relevant narratives is an expositional device to present some of our main ideas in a simple form, while deferring others to the next section.

Recall that under rational expectations, the only prevailing platform is $(g, \{1\}, \textit{true})$—that is, group 1 is always in power and it implements high taxation. The following result characterizes equilibrium when the two false narratives, *denial* and *tribal*, are also feasible.

PROPOSITION 1. *There is a unique equilibrium σ^*. The only platforms that can be in Supp(σ^*) are $(g, \{1\}, \textit{true})$, $(b, \{2\}, \textit{denial})$, and $(b, \{1\}, \textit{tribal})$. Furthermore,*

(i) $\sigma^*(g, \{1\}, \textit{true}) = \min\{1, f(1, g)/f(2, b)\}$;

(ii) $\sigma^*(b, \{1\}, \textit{tribal}) > 0$ *only if* $\sigma^*(b, \{2\}, \textit{denial}) > 0$.

The proofs of all the formal results are in the Appendix.

To interpret the equilibrium, assume $f(2, b) > f(1, b) > f(1, g) > f(2, g)$. This is a natural restriction given our public-good story, because it means that *ceteris paribus*, both groups find low taxation intrinsically more appealing. It also ensures that all three platforms mentioned in Proposition 1 are in *Supp*(σ). When *true* prevails, this means that group 1 is in power, implements high taxation, and employs the true narrative, which attributes outcomes to policies. This narrative essentially claims that high taxation leads to successful public-good provision. When *denial* prevails, this means that group 2 is in power, implements low taxation, and employs the denial narrative, which implicitly attributes the outcome to external factors such as technological changes. Finally, when *tribal* prevails, this means that group 1 is in power, implements low taxation, and employs the tribal narrative. This narrative credits one party for successful public-good provision, without being specific about policies. The three narratives roughly correspond to those described by Stone (1989), as quoted in the Introduction.

Our result generates endogenous fluctuations in the identity of ruling parties and the policies they implement, including policy shifts within a ruling party. In particular, the "big government party" sometimes implements the "small government" policy. In conventional political-economics models, such fluctuations would be attributed to changes in primitives, such as voters' preferences. In our model, they are an

equilibrium consequence of competition over public opinion, fueled by false narratives that misinterpret historical correlations between outcomes, policies and ruling parties.

To gain intuition for Proposition 1, let us write the expressions for the total mobilization generated by the three platforms:

$$\begin{aligned} M_\sigma(g,\{1\},\textit{true}) &= p_\sigma(y=G \mid a=g)\cdot f(1,g) = q\cdot f(1,g)\\ M_\sigma(b,\{2\},\textit{denial}) &= p_\sigma(y=G)\cdot f(2,b) = q\cdot p_\sigma(a=g)\cdot f(2,b)\\ M_\sigma(b,\{1\},\textit{tribal}) &= p_\sigma(y=G \mid x_1=1)\cdot f(1,b)\\ &= q\cdot p_\sigma(a=g \mid C=\{1\})\cdot f(1,b). \end{aligned}$$

In equilibrium, the rational, high-taxation policy g must occur with positive probability. The reason is that any platform carried by a false narrative free-rides on episodes of high taxation. Also, note that a platform advocating high taxation will generate its largest total mobilization if it employs the true narrative, which highlights the correlation between a and y (this correlation is stronger than the correlation between y and any other variable).

However, when $f(2,b) > f(1,g)$, the low taxation policy b has higher mobilization potential than g. False narratives generate wrong beliefs that allow b to gain dominance at the expense of g. They enable small-government supporters "eat their cake and have it". On the one hand, they are intrinsically attracted to low taxation. On the other hand, false narratives distract them from the adverse consequences of low taxation. The equilibrium probability of high taxation is determined by the ratio $f(1,g)/f(2,b)$. What makes low taxation not only popular but also "populist" is that it necessitates a false narrative to mobilize public opinion.

The distinction between the two false narratives—denial and tribal—is irrelevant for the equilibrium probability of $a=g$. However, it matters for the identity of the ruling party. When $f(1,b) > f(1,g)$, the tribal narrative enables group 1 to displace group 2, even though it adopts the same "populist" policy b. The reason is that group 1 can milk its reputation for achieving successful public-good provision—thanks to its tendency to implement high taxation. It does so by highlighting the long-run correlation between $y=G$ and being in power (or, equivalently, group 2 being out of power).

A Dynamic Interpretation. For a deeper intuition behind the equilibrium, it is useful to have a dynamic process in mind. At every time period, the mobilization value (or M-value) of platforms is calculated according to the historical frequencies of prevailing platforms; the platform with the highest M-value is the one that prevails at that period. Imagine that initially there are random fluctuations over (a,C) and only the true narrative is considered. This narrative can only justify policy g because $\Pr(y=G \mid a) = q\cdot \mathbf{1}[a=g]$. This policy mobilizes group 1 more strongly; hence, the prevailing platform is $(g,\{1\},\textit{true})$.

Suppose this status quo persists for a while, and at some point platform $(b,\{2\},\textit{denial})$ arises. By then, the historical frequency of $a=g$ is close to one.

Therefore, the denial narrative induces the belief $\Pr(y = G) \approx q$. Because $f(2, b) > f(1, g)$, the new platform is more strongly mobilizing than the "incumbent" platform $(g, \{1\}, \textit{true})$. As a result, the new platform displaces the old one and becomes dominant. Since the new platform involves policy b, the historical frequency of policy g gradually declines, lowering $\Pr(y = G)$.

As this process continues, the denial platform's mobilization drops below $q \cdot f(1, b)$. At that same time, the platform $(b, \{1\}, \textit{tribal})$ gains traction. In the path described so far, $a = g$ is strongly associated with $x_1 = 1$. This implies the historical conditional probability $\Pr(y = G \mid x_1 = 1) \approx q$. Consequently, a narrative arguing that things are good when group 1 is in power (or, equivalently, when group 2 is out of power) can mobilize group 1 behind policy b. The total mobilization of $(b, \{1\}, \textit{tribal})$ is approximately $q \cdot f(1, b)$. Since $f(1, b) > f(1, g)$, this exceeds the total mobilization of the two previous platforms, and $(b, \{1\}, \textit{tribal})$ becomes dominant. As this phase continues, it gradually weakens the correlation between x_1 and y and therefore lowers the total mobilization that the platform generates. By lowering the frequency of $y = G$, it also weakens the appeal of the denial narrative. This brings the platform carried by the true narrative back in vogue.

The subsequent dynamic repeats this cycle, albeit with smaller swings in total mobilization because marginal and conditional frequencies are calculated over longer histories. In the long run, all three platforms generate the same total mobilization $q \cdot f(1, g)$. Any deviation that raises the long-run frequency of one platform will trigger an offsetting dynamic response. That is, the equilibrium of Proposition 1 is dynamically stable. Section 6.2 formalizes this process in the context of the general multi-group case.

Comment. The assumption that $\Pr(y = G \mid b) = 0$ was made for tractability, as it enables the convenient multiplicative form of the equations that characterize equilibrium mobilization. We believe that our qualitative results would hold when $\Pr(y = G \mid b) < q$, as long as $\sum_i f(i, g) / \sum_i f(i, b)$ is not too low. This would ensure that g is implemented with positive probability in equilibrium, which, in turn, would anchor the equilibrium mobilization level, as in the current analysis.

4. Fragmented Societies

This section considers societies with more than two social groups ($n > 2$). Relative to Section 3, three key differences emerge. First, "*exclusionary*" narratives of the form "things are good when these groups are *out* of power" are no longer equivalent to "*inclusionary*" narratives of the form "things are good when these groups are *in* power". We will see that only the former arise in equilibrium. Second, the proliferation of exclusionary narratives can depress the equilibrium probability of the good outcome. Finally, new coalition structures can arise that would not be sustainable if only the true narrative were feasible.

An Example with a Fragmented Left. For concreteness, suppose that the issue is national security. Let g and b represent hawkish and dovish policies, and let G and B represent good and bad national-security outcomes. Thus, in this example, hawkish policy is necessary (but not sufficient) for a good national-security outcome. There are four social groups ($n = 4$), classified as follows: The "Right" is $\{1\}$, the "Center" is $\{2\}$, and the "Left" is $\{3, 4\}$. The domain of feasible narratives is $\{\{1\}, \{2\}, \{3\}, \{4\}, \{3, 4\}\}$.

Social groups' mobilization propensities reflect ideological attitudes to national-security policies. Specifically, $f(1, b) = f(3, g) = f(4, g) = 0$ — that is, the Right (Left) is ideologically opposed to b (g). In addition, assume that $f(2, b) > f(1, g) + f(2, g)$ — that is, the Center's mobilization propensity given b is stronger than the mobilization propensity given g among the Center–Right. The interpretation is that the Center is non-ideological and therefore does not oppose any policy; it finds the dovish policy intrinsically more appealing because it requires fewer sacrifices than the hawkish policy. Finally, to make calculations easier to follow, let $f(3, a) \equiv f(4, a)$.[10]

The following distribution is an equilibrium (indeed, the unique one in a sense we will make precise below):

σ	*policy*	*coalition*	*narrative*
$\frac{f(1,g)+f(2,g)}{f(2,b)+f(3,b)+f(4,b)}$	g	$\{1, 2\}$	*true*
$\frac{f(2,b)-f(1,g)-f(2,g)}{f(2,b)+f(3,b)+f(4,b)}$	b	$\{2\}$	$\{3, 4\}$
$\frac{f(3,b)+f(4,b)}{2[f(2,b)+f(3,b)+f(4,b)]}$	b	$\{2, 3\}$	$\{4\}$
ditto	b	$\{2, 4\}$	$\{3\}$.

As in two-group societies, the dovish policy b occurs with positive probability and it is sustained by false tribal narratives that take an "exclusionary" form. For example, in platform $(b, \{2\}, \{3, 4\})$, the Center attributes a good national-security outcome to keeping the Left out of power. Furthermore, the equilibrium exhibits *endogenous fragmentation*: Each faction of the Left sometimes joins the Center to form a coalition, using a false narrative that attributes the good outcome to keeping the remaining left-wing group out of power. As we will see, this is a general feature of equilibrium in the multi-group model.

The equilibrium exhibits a coalitional configuration we sometimes observe in multi-party political systems. A pragmatic "centrist" group has a stable hold on political power (its pragmatism consists of adopting both policies in equilibrium). It is sometimes joined by ideologically pure groups on either side of the isle. Our model interprets this pattern as an equilibrium consequence of public-opinion politics, fueled by false narratives.

10. It would be more natural to assume that $f(3, \cdot)$ and $f(4, \cdot)$ are different, reflecting ideological subdivisions within the Left. As it stands, our specification is akin to the distinction between the "Judean People's Front" and the "People's Front of Judea".

The example has two additional noteworthy features. First, the equilibrium probability of the rational policy g is equal to $\sum_i f(i,g)/\sum_i f(i,b)$, the ratio between the two policies' total mobilization propensity. Second, the entire Left category $\{3,4\}$ is invoked by one of the prevailing tribal narratives, yet ruling coalitions never contain it. Thus, a political entity can be relevant for tribal narratives even if it never belongs to a ruling coalition in its totality. For example, a right-wing party can use a scapegoating narrative that invokes "the Left", lumping together the moderate Left and radical Left, even if the two are never in the same government. □

To proceed with the general analysis, let $N^a = \{i \in N \mid f(i,a) > 0\}$ denote the set of social groups that do not oppose policy a. For convenience, we will refer to $N \setminus N^b$ as the "Right", $N \setminus N^g$ as the "Left," and $N^g \cap N^b$ as the "Center". For every feasible narrative S, let $L(S)$ be the components of S that belong to the Left:

$$L(S) \equiv S \cap (N \setminus N^g).$$

For every $J \subseteq N$, let $F(J,a)$ be the aggregate mobilization propensity given a of the groups in J:

$$F(J,a) \equiv \sum_{i \in J} f(i,a).$$

When $F(N,g) > F(N,b)$ that is, when the population finds g more appealing than b—it follows immediately from (3) to (4) that $M_\sigma(g, N^g, \{0\}) > M_\sigma(b, C, S)$ for every C, S. In this case, $\Pr(a = g) = 1$ in any equilibrium. Moreover, $M_\sigma(g, N^g, \{0\}) \geq M_\sigma(g, C, S)$ for every C, S, and thus there is an equilibrium σ in which $\sigma(g, N^g, \{0\}) = 1$.

The next result provides an equilibrium characterization for the more interesting case in which $F(N,g) \leq F(N,b)$. The proof develops an algorithm to compute the unique equilibrium distribution over (a, C).

THEOREM 1. *Let $F(N,g) \leq F(N,b)$. An equilibrium σ^* exists. Furthermore, any equilibrium induces the same unique distribution over policy-coalition pairs (a, C) and has the following additional properties:*

(i) *The policy g is played with positive probability which is at most $F(N,g)/F(N,b)$.*
(ii) *If $(g, C, S) \in Supp(\sigma^*)$, then $C = N^g$ and $0 \in S$.*
(iii) *Every platform $(b, C, S) \in Supp(\sigma^*)$ satisfies $S \subseteq N^b$ and $C = N^b \setminus L(S)$.*

The first part of this result establishes that the equilibrium probability of the rational policy is positive. It also provides an upper bound on this probability. The bound is implied by the denial narrative, in the following sense. The total mobilization generated by $(g, N^g, \{0\})$ is $q \cdot F(N,g)$, which in equilibrium has to be weakly larger than the total mobilization generated by $(b, N^b, \varnothing)$, namely $q \cdot p_{\sigma^*}(a = g) \cdot F(N,b)$. This inequality implies the upper bound.

Theorem 1 only partially pins down equilibrium narratives. The reason is that multiple narratives can induce the same belief, and therefore the same total mobilization. In particular, if $0 \in S$, then $p_\sigma(y \mid x_S(a, C)) = p_\sigma(y \mid a)$ because y is independent of C conditional on a (as we saw in Section 2.1).

Therefore, it is convenient to focus on equilibria in which narratives do not have any redundant component.

DEFINITION 3 (*Essential equilibria*). An equilibrium σ is essential if whenever $(a, C, S) \in Supp(\sigma)$, then: (i) if $p_\sigma(y \mid a) = p_\sigma(y \mid x_S(a, C))$ for all a, C, then $S = \{0\}$; and (ii) there is no $T \subset S$ such that $p_\sigma(y \mid x_T(a, C)) = p_\sigma(y \mid x_S(a, C))$ for all a, C.

This refinement applies two "tie-breaking rules" that favor the true narrative over false ones, and small narratives over large ones. This enables us to obtain a sharper characterization of equilibrium narratives, under a mild restriction of the domain of feasible narratives.

COROLLARY 1. *Suppose that if S is a feasible narrative, then $S \setminus (N^g \cap N^b)$ is also feasible. Then, there exists a unique essential equilibrium σ^*. Furthermore,* (i) *if $(g, C, S) \in Supp(\sigma^*)$, then $S = \{0\}$ and $C = N^g$; and* (ii) *if $(b, C, S) \in Supp(\sigma^*)$, then $S \subseteq N \setminus N^g$ and $C = N^g \setminus S$.*

Thus, in the unique essential equilibrium, the rational policy g is accompanied by the true narrative, whereas the false narratives that accompany policy b take the exclusionary tribal form. They identify a collection S of groups that oppose g, but are not in the coalition supporting b. By attributing the outcome to the power status of S, the narrative essentially argues that "things are good when S is out of power". The denial narrative is a special case in which $S = \varnothing$. When this narrative is employed, the ruling coalition consists of the Center and the entire Left.

Corollary 1 shows that exclusionary and inclusionary tribal narratives are no longer equivalent when $n > 2$. What makes exclusionary narratives more effective? When a group opposes g, there is positive correlation between that group being out of power and the good outcome. The exclusionary narrative exploits this correlation to generate a false belief that the very exclusion of specific groups from power will lead to a good outcome, while advocating policy b. This enables groups to "*have their cake and eat it*:" They reap the mobilization benefits of the intrinsically more attractive b, while deflecting responsibility for a bad outcome and "scapegoating" the excluded groups for it.

Incontrast, platforms advocating b refrain from using "inclusionary" narratives that attribute the outcome to the power status of coalition members. To gain intuition, suppose that a platform advocating b employs a narrative that includes groups $inside$ the platform's coalition. For the narrative to be effective, the presence of these groups in ruling coalitions should be *positively* correlated with the good outcome. This means these groups must support both b and g—that is, they are "centrists". Moreover, these

groups would never be scapegoated, because their exclusion from ruling coalitions is *negatively* correlated with the good outcome. Therefore, in equilibrium, these groups would join every ruling coalition—that is, they would *always* be in power. This equilibrium effect means that these groups' power status is uncorrelated with the outcome, thus making the inclusionary tribal narrative ineffective (it has no advantage relative to the denial narrative).

Both inclusionary and exclusionary tribal narratives S are "simple" in the sense that they point to social groups with identical power status—that is, either all of them are in the coalition C or none of them is. In principle, one could have tribal narratives S that are "hybrid" with respect to C—for example,$S = \{1, 2\}$, $1 \in C$ and $2 \notin C$. The characterization in Theorem 1 allows for such narratives, whereas Corollary 1 rules them out—although with no substantive consequence as clarified by the definition of essential equilibrium.

Exclusionary tribal narratives trade off breadth and intensity of political mobilization. Excluding groups from a coalition is costly because it forgoes their mobilization propensity. However, if this exclusion is not too frequent, its correlation with $a = g$ (and hence $y = G$) remains strong, thus generating intense support from the coalition members. At one extreme, the denial narrative garners the largest coalition by not excluding any group, but induces a weaker belief of $y = G$ by not exploiting any correlation in the data.

Tribal narratives give rise to coalitions that would be impossible otherwise. If the true and denial narratives were the only feasible ones, the equilibrium support would not feature coalitions other than N^g and N^b. Thanks to tribal narratives, strict subsets of N^b appear as equilibrium coalitions.

The following result characterizes when *non-empty* exclusionary narratives are part of the unique essential equilibrium.

PROPOSITION 2. *There exists* (b, C, S) *with non-empty* $S \subset N$ *in the support of the essential equilibrium if and only if* $0 < F(T) < F(N, b) - F(N, g)$ *for some feasible narrative* $T \subseteq N \setminus N^g$.

The condition is that the domain of feasible narratives induces a set whose aggregate mobilization propensity is sufficiently weak—and so it is not too politically costly to exclude. When the condition is violated, the only false narrative that can be part of essential equilibrium is the denial narrative.

5. Specific Domains of Feasible Narratives

Section 4 allowed for any domain of feasible narratives that includes the true and denial narratives. In this section, we consider various restricted domains. We use $\mathcal{S}$ to denote the domain of feasible *tribal* narratives (i.e., $S \subseteq N$ for every $S \in \mathcal{S}$). There are several reasons for considering such restricted domains. First, we interpret each $S \in \mathcal{S}$ as a collection of social groups that can be *clearly identified* by a common label or

defining attribute ("fundamentalists", "progressive left", "unionized workers" or "the economic elite"). Second, $\mathcal{S}$ reflects the extent to which different social groups have distinct political representation, which can render them accountable for outcomes. In some political systems (e.g., Israel), there are political parties that directly represent specific ethno-religious groups. Consequently, there is data about their power status and how it is correlated with outcomes, which makes a narrative that exploits this correlation quantifiable. In other systems (e.g., the US), the mapping between specific social groups and political representation is more blurred, thus restricting the supply of similar narratives.

This section is structured as follows. In Section 5.1, we consider a particular restricted domain and show that it leads to a simple characterization of $\Pr(a = g)$ and equilibrium narratives. Section 5.2 characterizes the narrative domains for which $\Pr(a = g)$ hits the upper bound provided by Theorem 1. Section 5.3 applies this characterization to other specific domains.

Throughout the section, we assume that policy b is intrinsically more appealing than policy g, even among the groups that intrinsically support g. That is, mobilization propensity satisfies

$$F(N^g, b) > F(N^g, g). \tag{5}$$

This condition fits situations in which g is a more costly policy (carbon tax, fiscal restraint) and therefore, *ceteris paribus*, it is intrinsically less popular than b. For expositional convenience, this section focuses on *essential equilibria* (as defined and characterized in Section 4).

5.1. A Multi-Attribute Model

Suppose that each social group is characterized by multiple attributes that represent ideological, ethno-religious, or socioeconomic identities. That is, let $N = \{0,1\}^K$, where $K > 1$.[11] Use $i_k \in \{0,1\}$ to denote the value of group i's kth attribute, and denote $i_B = (i_k)_{k \in B}$.

Let $m \in \{0, \ldots, K-1\}$ and assume that $N \setminus N^g = \{i \in N \mid i_k = 1 \text{ for all } k > m\}$. That is, specific values of the attributes $m+1, \ldots, K$ identify the Left category. The set of groups on the Left are effectively defined by $\{0,1\}^{1,\ldots,K}$, such that m indicates the degree of *internal fragmentation* among the Left.

Suppose $\mathcal{S}$ contains all sets $S \subset N$ that take the form $S = \{i \in N \mid i_B = v\}$ for some $B \subseteq \{1, \ldots, K\}$ and $v \in \{0,1\}^B$. That is, a feasible tribal narrative focuses on some subset of attributes B and fixes their values; the narrative is defined as the set of groups that share these values. For example, $S = \{i \in N \mid i_1 = 1, i_2 = 0\}$ is a feasible narrative. For example, in the context of Israeli politics, it can represent a narrative that attributes outcomes to the power status of religious Jews. In contrast, $S = \{i \in N \mid i_1 = i_2\}$ is not a feasible narrative in this multi-attribute model.

11. The restriction to *binary* attributes is for expositional simplicity; the analysis easily extends to an arbitrary finite alphabet.

PROPOSITION 3. *In the unique essential equilibrium σ^* of the multi-attribute model,*

$$p_{\sigma^*}(a = g) = \frac{F(N, g)}{F(N^g, b) + \max\{m, 1\} \cdot F(N \setminus N^g, b)}. \tag{6}$$

Furthermore, the narratives that accompany $a = b$ in the support of σ^* are $S = N \setminus N^g$ and all sets of the form

$$S = (N \setminus N^g) \cap \{i \in N \mid i_k = v\}, \tag{7}$$

for some $k \in \{1, \ldots, m\}$ and $v \in \{0, 1\}$.[12]

This result has two noteworthy features. First, the exclusionary tribal narratives that sustain policy b in equilibrium take a simple form. One such narrative is $S = N \setminus N^g$. The coalition that accompanies this combination of a and S is the Center $C = N^g \cap N^b$—that is, in this platform, the Center scapegoats the entire Left. The other narratives that accompany policy b scapegoat all Left groups having a particular value $v \in \{0, 1\}$ in one of the attributes $k \in \{1, \ldots, m\}$ that distinguish among them. For example, suppose attribute $k \leq m$ indicates a social group's education status. Then, one of the equilibrium narratives that accompany policy b can be phrased as "the outcome is good when the highly educated Left is out of power".

Second, expression (6) gives an explicit formula for the equilibrium probability of policy g. This probability decreases with m (strictly so when $m > 1$). Thus, political fragmentation on the Left creates more room for false tribal narratives that crowd out the true narrative and the rational policy g.

The formula suggests an additional comparative-statics exercise. Consider changes in mobilization propensities that reflect *more polarized attitudes* toward policy b. Specifically, suppose $F'(N^g, b) = F(N^g, b) - \varepsilon$ and $F'(N \setminus N^g, b) = F(N \setminus N^g, b) + \varepsilon$, where $\varepsilon > 0$ is small enough that condition (5) continues to hold. This change from F to F' captures a shift of intrinsic support for b from the Center to the Left, resulting in a more polarized society. When $m > 1$, this shift lowers $p_{\sigma^*}(a = g)$. In this sense, higher polarization is detrimental to the rational policy.

5.2. When do Tribal Narratives Crowd out Rational Policies?

We now characterize the tribal-narrative domains $\mathcal{S}$ for which the equilibrium probability of policy g achieves the upper bound $F(N, g)/F(N, b)$. Recall that this bound is attained when denial is the only feasible false narrative. Therefore, when the equilibrium probability of $a = g$ hits the upper bound, it means that tribal narratives are policy-irrelevant.

12. We will prove this result by applying the general characterization theorem presented in the next subsection.

We say that $S \subset N \setminus N^g$ is a *coarse subcategory* of the Left if there is no S' such that $S \subset S' \subset N \setminus N^g$ (It is understood that both S and S' are in $\mathcal{S}$.) We also introduce the following properties of $\mathcal{S}$:

(i) $S \cup \widehat{S} = N \setminus N^g$ for all coarse subcategories S and $\widehat{S}$ of the Left.

(ii) For every $S \in \mathcal{S}$, $S \subset N \setminus N^g$, that is not a coarse subcategory of the Left,

$$S = \bigcap_{S \subset S'} S'.$$

Property (i) says that coarse subcategories are sufficiently broad so that every pair of them covers the Left. Property (ii) says that every finer category is equal to the intersection of its coarser categories.

THEOREM 2. *Fix $F(N^g, b)$ and $F(N, g)$ (and recall that $F(N^g, b) > F(N, g)$). Then, in any equilibrium $\sigma^*, p_{\sigma^*}(a = g) = F(N, g)/F(N, b)$ for all values of $F(N \setminus N^g, b)$ if and only if $\mathcal{S}$ satisfies properties* (i) *and* (ii).

This result says that exclusionary tribal narratives cannot crowd out the rational policy—no matter how strongly the Left supports b—if and only if properties (i) and (ii) hold. To illustrate the result, reconsider the multi-attribute model. Coarse subcategories in this model are obtained by fixing the value of one attribute $k \leq m$. For example, suppose S and S' correspond to fixing $i_m = 1$ and $i_{m-1} = 1$. Then, $S \cup S' = \{i \in N | i_m = 1 \text{ or } i_{m-1} = 1\}$, which is a strict subset of $N \setminus N^g$. It follows that property (i) fails that is why $p_{\sigma^*}(a = g) < F(N, g)/F(N, b)$.

It is easy to verify that the multi-attribute model does satisfy property (ii). Lemma A.1 in the proof of Theorem 2 establishes that property (ii) is necessary and sufficient for the feature that coarse subcategories of the Left are the smallest tribal narratives that are employed in every essential equilibrium. This is indeed the case in the equilibrium given by Proposition 3. The next sub-section further illustrates the role that properties (i) and (ii) play in the characterization of essential equilibrium.

5.3. Additional Examples of Narrative Domains

A Hierarchical Multi-Attribute Model. The multi-attribute model assumes that a feasible narrative is defined by setting the values of some collection of attributes B. However, in some applications, we may wish to impose additional structure. For example, the attributes may be *hierarchically ordered*, such that the distinction between values of attribute k is nonsensical unless the value of attribute $k + 1$ has been pinned down. For example, attribute $k + 1$ may indicate social groups' broad religious identity (e.g., Jewish), while attribute k indicates their finer religious affiliation (e.g., Orthodox). Therefore, a narrative that specifies the value of attribute k must also specify the value of attribute $k + 1$.

To capture this idea, let $D \in \{1, \ldots, m\}$ be a constant, and define $\mathcal{S}$ as the collection of all $S \subset N$ that take the form $S = \{i \in N \mid i_{\{k,\ldots,K\}} = v\}$ for some $k \in$

$\{m-D+1,\ldots,K\}$ and $v \in \{0,1\}^{\{k,\ldots,K\}}$. This specification represents a "*social taxonomy*": The narrative defined by $v_k,\ldots,v_K$ is a direct subcategory of the coarser category defined by $v_{k+1},\ldots,v_K$. The parameter D represents the depth of the social taxonomy.

PROPOSITION 4. *In the hierarchical multi-attribute model, the unique essential equilibrium σ^* satisfies*

$$p_{\sigma^*}(a=g) = \frac{F(N,g)}{F(N^g,b) + D \cdot F(N \setminus N^g,b)}. \tag{8}$$

This formula is similar to (6), except that D replaces m. Note that $p_{\sigma^*}(a=g) < F(N,g)/F(N,b)$ if and only if $D>1$. In fact, the hierarchical multi-attribute model violates property (ii) unless $D=1$—because the intersection of narratives coarser than S is the smallest S' that strictly contains S. However, property (i) holds because coarse subcategories of the Left partition $N \setminus N^g$ into two subsets pinned down by the value of attribute $m-1$.

The structure of equilibrium narratives is qualitatively different between the hierarchical and the non-hierarchical (original) multi-attribute model. In the latter, only a fraction of the feasible tribal narratives are employed in equilibrium. In contrast, in the hierarchical model, *every* feasible narrative $S \subseteq N \setminus N^g$ is realized with positive probability in the essential equilibrium. To see why, suppose an exclusionary tribal narrative invokes some category S' in the social taxonomy, and yet one of its direct sub-categories S is never invoked. The hierarchical structure of $\mathcal{S}$ implies that the equilibrium narratives that weakly contain S' and S are the same. This means that narratives S and S' generate the same beliefs. However, the smaller S is coupled with a larger coalition and therefore generates higher total mobilization than does S', so we cannot be in an equilibrium.

A Rich Domain of Tribal Narratives. Finally, consider the extreme case in which $\mathcal{S}$ is the set of *all* subsets $S \subseteq N$. We refer to such $\mathcal{S}$ as the "rich" narrative domain. The multi-attribute structure of N is redundant in this case, so we ignore it here.

PROPOSITION 5. *In the unique essential equilibrium σ^* under a rich narrative domain, $p_{\sigma^*}(a=g) = F(N,g)/F(N,b)$. Furthermore, the narratives that accompany policy b in the support of the equilibrium are $S = N \setminus N^g$ and all sets of the form $S = N \setminus (N^g \cup \{i\})$ for some $i \in N \setminus N^g$.*

The proof of this result is a simple application of Theorem 2. The rich domain satisfies both properties (i) and (ii). Property (i) holds because coarse subcategories of the Left correspond to $N \setminus (N^g \cup \{i\})$ for any $i \in N \setminus N^g$. Property (ii) holds because any intersection of subsets of $N \setminus N^g$ is by definition in $\mathcal{S}$. Therefore, the equilibrium probability of $a=g$ attains the upper bound in Theorem 1. Turning to the structure of equilibrium false narratives, $N \setminus N^g$ and its coarse subcategories are employed as exclusionary tribal narratives; the proof is exactly as in the case of

Proposition 3. Since the rich domain satisfies property (ii), Lemma A.1 implies that these are the only false narratives that are employed in equilibrium. Thus, the narratives that accompany policy b take the following form: Either the entire Left $N \setminus N^g$ is scapegoated or the Left minus exactly one group is scapegoated (this group joins the Center to form a Center–Left ruling coalition).

Proposition 5 demonstrates that the effect of political fragmentation on $p_{\sigma^*}(a = g)$ is nonmonotonic. The rich domain represents a larger scope for tribal narratives than the multi-attribute domain. Nevertheless, $p_{\sigma^*}(a = g)$ is higher whenever $m > 1$. The reason is that apart from narrative $N \setminus N^g$, which belongs to both domains, the largest narratives in the rich domain are larger than the largest narratives in the multi-attribute domain. This means that the coalitions that employ false narratives tend to be smaller in the rich domain case, which is compensated for by a more optimistic belief, namely, a larger $p_{\sigma^*}(a = g)$.

To summarize our findings for the three domain restrictions, we considered the rich domain and multi-attribute domain are similar in the equilibrium structure of false narratives, but differ in terms of the equilibrium probability of policy g. In contrast, the hierarchical and non-hierarchical multi-attribute domains are similar in the equilibrium probability of g (in terms of the mobilization propensity function and the measure of political fragmentation), but differ in the structure of equilibrium narratives.

Comment: The Cohesiveness of Political Scapegoats. In the example of Section 4 and the restricted domains examined in this section, the denial narrative never features in equilibrium. Furthermore, according to Theorem 2, this is a necessary equilibrium property whenever $\mathcal{S}$ violates properties (i) or (ii) (because if denial is employed, $\Pr(a = g)$ hits the upper bound). In these cases, the Left $N \setminus N^g$ never belongs as a whole to a ruling coalition in equilibrium, even though it features as a tribal narrative when the ruling coalition is the Center $N^g \cap N^b$. What makes "the Left" a cohesive political entity is that it sometimes belongs as a whole to the *opposition*. For this reason, the exclusionary tribal narrative $S = N \setminus N^g$ (which amounts to saying that "things are good when the Left is not in power") is observationally meaningful.

6. Foundations

In this section, we provide "micro-foundations" for our notion of political mobilization given by Definition 1, and a dynamic foundation for our equilibrium concept given by Definition 2.

6.1. Micro-foundations for the Mobilization Rule

In Section 2, we remarked that the functions f and m underlying our model of political mobilization can be "justified" as arising from of social-group members' underlying preferences. In this sub section, we provide two alternative "micro-foundations" that formalize this claim.

The first micro-foundation is a variant of what Persson and Tabellini (2000) refer to as a "simple model of public finance", alluded to in Section 3 (see Ch. 3.1 in their book). Suppose the two policies b and g represent "small government" and "big government". The outcomes G and B represent successful and failed provision of a public good and generate gross material payoffs of 1 and 0 for every individual in society, respectively. Each policy a is associated with a tax rate τ_a, where $\tau_b < \tau_g$. Each social group i consists of a measure one of individuals, whose gross income is uniformly distributed over the interval $[0, d_i]$.

Consider an individual with income x, who believes that a platform that includes the policy a delivers successful prevision of the public good with probability α. The individual's net anticipatory payoff from the platform is $\alpha - \tau_a x$. In addition to his income, the individual is characterized by whether he ideologically approves of each policy. The individual supports the platform if and only if he ideologically approves of its policy and if it delivers him a positive net anticipatory payoff. Let $\gamma(i,a)$ denote the fraction of individuals in group i who ideologically approve of a. Then, for suitable values of τ and d, the total measure of individuals from group i in support of the platform is $\gamma(i,a)\alpha/\tau_a d_i$. This specification fits our definition of m, for $f(i,a) = \gamma(i,a)/\tau_a d_i$.

The second micro-foundation is based on the interpretation that the policy b produces immediate results, whereas the benefits of the policy g are realized with delay. In other words, the two policies represent short- and long-term measures. As a result, the outcome of g needs to be discounted, unlike the outcome of b. Let $\delta_i(a)$ denote the discount rate that social group i applies to policy a, where $0 \leq \delta_i(g) < \delta_i(b) = 1$ for each i. For each group, the undiscounted payoffs from the outcomes G and B are 1 and 0, respectively. Objectively, only the long-term measure can bring a good outcome. False narratives can induce the belief that the short-term measure can be successful.

Thus, when group i believes that a platform, which includes the policy a delivers the outcome $y = G$ with probability α, the group's anticipatory utility from the platform is $\delta_i(a)\alpha$. Suppose that for each group, the members' cost of political participation is uniformly distributed over [0,1]. An individual is mobilized to support a platform whenever the anticipatory utility he derives from it exceeds his participation cost. Let s_i denote the *size* of group i. Therefore, the strength of a group's support for a platform is $s_i\delta_i(a)\alpha$. This is a micro-foundation of our definition of m, for $f(i,a) = s_i\delta_i(a)$. It has the additional structure that $f(i,g) \leq f(i,b)$ for every i—that is, all groups find the action b intrinsically more attractive. The micro-foundation imposes no additional restrictions.

6.2. A Dynamic Foundation for the Equilibrium Concept

In this sub section, we consider a simple and natural dynamic process that determines which platforms garner maximal popular support over time. We show that the process converges to the unique equilibrium distribution over policies and coalitions in our main result (Theorem 1). This global convergence result provides a dynamic foundation for our equilibrium concept.

Time is discrete and denoted by $t = 1, 2, \ldots$. In each period t, there is a distribution σ_t over platforms (a, C, S), where $a \in \{b, g\}$, $C \subseteq N$, and $S \in \mathcal{S}$. Let the initial σ_1 be any distribution with full support over the set of platforms using admissible coalitions. Since the set of platforms is finite, this distribution is well-defined. The distribution σ_t evolves according to the following adjustment. For every $t \geq 2$, let

$$\overline{(a, C, S)}_t \in \arg\max_{(a', C', S')} M_{\sigma_t}(a', C', S'),$$

where ties can be broken arbitrarily. Then, let

$$\sigma_{t+1}(a, C, S) = \begin{cases} \frac{1}{t+1} + \frac{t}{t+1}\sigma_t(a, C, S) & \text{if } (a, C, S) = \overline{(a, C, S)}_t \\ \frac{t}{t+1}\sigma_t(a, C, S) & \text{otherwise.} \end{cases}$$

Thus, for t large enough, we can essentially view $\sigma_t(a, S, C)$ as the empirical frequency with which platform (a, C, S) has been dominant in the available history of data.

PROPOSITION 6. *Every limit point σ of the process σ_t induces the same distribution over policy-coalition pairs (a, C) as that induced by the unique essential equilibrium σ^*.*

This result formalizes and generalizes the dynamic convergence process we discussed in the context of the two-group specification in Section 3.

7. The Dissociation between Policies and Outcomes

The ability to dissociate intrinsically appealing policies from their bad outcomes is the key to the thriving of false narratives in our model. This type of dissociation is common in everyday life, as a psychological mechanism for dealing with cognitive dissonance. For example, think of a worker who prefers not to work hard. Faced with the resulting poor workplace outcomes, he "gets himself off the hook" by attributing these outcomes to the quality of his bosses.

Our model takes this sort of self-deception from the personal to the political domain. The public-good story invoked in Sections 3 and 6.1 offers a concrete example. The inflation debate discussed in the Introduction is another case in point. The outcomes G and B represent low and high inflation, and the policies g and b stand for fiscal restraint and fiscal expansion. In this context, arguments that we "live in a post-inflation age", or that inflation is a consequence of supply shocks or corporate greed are real-life analogues of our denial narrative. They give governments a license to run budget deficits by denying their inflationary implications. Likewise, Paul Krugman's quote from the Introduction is akin to our tribal narrative, which attributes inflation to the party in power without explicating the role of policy.

Debates over national security offer yet another real-life example. Consider a politician who argues that historically, national security has been in good shape when his party was in power. Even if this claim is factually true, it may be a consequence of costly past policy choices (high defense spending, territorial concessions). The politician's narrative enables him to invoke the historical correlation without requiring the costly implementation of these policies.

When politicians and public-opinion makers promote a policy-narrative bundle, they need not mention them in the same breath, as this might draw attention to the link between policies and outcomes. Instead, they can emphasize different aspects of the platform on different public occasions. One speech or social-media post will focus on policies, while another will highlight outcomes' spurious causes (i.e., the narrative). They can also deflect interviewers' demands that they acknowledge the contradictions between the two aspects.

Finally, the dissociation between policies and outcomes implied by tribal narratives offers a novel, critical perspective into *retrospective voting* [see Healy and Malhotra (2013) for a review article; Plescia and Kritzinger (2017) extend the concept to multi-party systems]. This is the notion that voters punish or reward parties according to their performance when they were in office. This view puts less emphasis on the policies that ruling parties take and more emphasis on outcomes. A conventional view is that retrospective voting improves government accountability and helps select competent candidates. Our view is that attributing public outcomes to who is (or is not) in power rather than to the implemented policies can be a false narrative that sustains policies with bad public outcomes.

8. Related Literature

Eliaz and Spiegler (2020) pioneered the formalization of political narratives as causal models, whose adoption by agents is driven by the (potentially false) prospective beliefs these models generate. The present paper borrows these basic ingredients and incorporates them into a new political-economics framework, offering a number of modeling innovations and asking fundamentally new questions. In contrast to Eliaz and Spiegler (2020), this paper considers a heterogeneous society and is the first to explore how false narratives serve as the "glue" of social coalitions and shape their structure. Furthermore, this paper investigates a new question of whether successful narratives attribute outcomes to what ruling parties do or to who they are—as in "tribal" narratives that emerge from our analysis. Finally, another novel contribution of this paper is to study the role of narratives in the link between political fragmentation and the quality of public policies.

More broadly, this paper is related to a strand in the political-economics literature that studies voters' belief formation according to misspecified subjective models or wrong causal attribution rules [e.g., Roemer (1994), Spiegler (2013), Esponda and Pouzo (2017), Izzo, Martin, and Callander (2021), Levy, Razin, and Young (2022), and Szeidl and Szűcs (2022)]. Among these, Roemer (1994) models voters who have

misspecified beliefs about the mapping from policies to outcomes. He shows that in such a setting parties may promote different views of how the economy works for strategic electoral reasons.

Levy, Razin, and Young (2022) studies dynamic electoral competition between two candidates, each associated with a different subjective model of how two policy variables map into outcomes. One model is complete and correct; the other is a "simplistic" model that omits one of the policy variables. Voter participation is costly; stronger beliefs lead to larger voter turnout. The long-run behavior of this system involves ebbs and flows in the relative popularity of the two models, not unlike the dynamics of platform popularity that underlie equilibrium in our model (see Section 6.2).

Finally, Szeidl and Szűcs (2022) analyze a model where a politician can persuade voters of a false alternative "reality" in which some given elite conspires to attack him. They show how this persuasion strategy can help the politician increase voters' support, limit his accountability, and spread distrust in elites outside the political domain.[13]

The general program of studying the behavioral implications of misspecified causal models is due to Spiegler (2016, 2020). In their general form, causal models are formalized as DAGs, following the Statistics/AI literature on graphical probabilistic models [Cowell et al. (1999), Pearl (2009)]. The causal models in this paper fit into the graphical formalism, but do not require its heavy use because they take a simple form: A clique that consists of the nodes that represent the outcome and the narrative variables (as well as the action variable as an isolated node, when it is not part of the narrative). This form is related to the misspecified models in otherwise very different works, such as Jehiel (2005), Eyster and Piccione (2013), or Mailath and Samuelson (2020)). Therefore, in this paper, graphical representations of causal models remained in the background.

Given the fluidity of the notion of narratives, it naturally invites diverse formalizations. Bénabou, Falk, and Tirole (2018) focus on moral decision-making and formalize narratives as messages or signals that can affect decision-makers' beliefs regarding the externality of their policies. Levy and Razin (2021) use the term to describe game-theoretic information structures that people postulate in order to explain observed behavior. Schwartzstein and Sunderam (2021a,b) propose an alternative approach to "persuasion by models", where models are formalized as likelihood functions and the criterion for selecting models is their success in accounting for historical observations. Shiller (2017) focuses on the spread of economic narratives in society, using an epidemiological analogy.

Our model involves competition between models (some of which are misspecified). The public selects between these models according to a criterion that reflects motivated reasoning. Cho and Kasa (2015) and Ba (2023) offer dynamic analyses of competing

13. For a survey on the broader field of behavioral political economy, see Schnellenbach and Schubert (2015).

models, when the selecting criteria involve empirical misspecification tests. Montiel Olea et al. (2022) study competition between models in the context of experts who vie for the right to make predictions.

The political science literature has long acknowledged the power of narratives in garnering public support for policies and in mobilizing people to protests or rallies [see Polletta (2008)]. In particular, the so-called "narrative policy framework" was developed as a systematic empirical framework for studying the role of stories or narratives in public policy. Studies employing this framework have argued that narratives have a greater influence on the opinions of policymakers and citizens than does scientific information [see the papers mentioned in the Introduction, or Jones, McBeth, and Shanahan (2014)].

Finally, there are a few recent attempts to study political and economic narratives empirically, using textual analysis. Mobilizing public opinion often takes the form of texts (speeches, op-eds, tweets). What we observe in these texts are qualitative stories more than bare quantitative beliefs. Ash, Gauthier, and Widmer (2021), Andre et al. (2022), and Macaulay (2022) have performed manual and machine analysis of these texts in order to elicit prevailing narratives in various contexts. Ambuehl and Thysen (2023) and Charles and Kendall (2023) used experimental methodology to shed light on the source of causal narratives' appeal.

9. Conclusion

This paper explored the role of false narratives in the mobilization of public opinion in heterogeneous societies. Our analysis gave rise to three main qualitative insights.

First, false narratives enable social groups to dissociate policies' intrinsic private appeal from their unattractive public outcomes, and thus enhance support for such policies. False narratives achieve this by attributing outcomes to spurious causes, exploiting historical correlations and misrepresenting them as causal.

Second, the false narratives that survive in equilibrium generally take an "exclusionary tribal" form (akin to "scapegoating"), arguing that keeping certain social groups *out of power* leads to good outcomes. The reason such narratives prevail is that they are consistent with a stable correlation between groups' power status and public outcomes. In contrast, "inclusionary" tribal narratives (which attribute outcomes to who is *in power*) are unstable because they effectively invite every group that supports the platform's policy to join the coalition. Consequently, the groups the narrative invokes are always in power, eroding any correlation that might give the narrative its apparent explanatory power in the first place.

Finally, political fragmentation leads to a proliferation of tribal narratives, which can exacerbate the underprovision of policies that deliver good public outcomes.

Appendix: Proofs

A.1. Proof of Proposition 1

We begin by recalling the total mobilization of platforms carried by the three relevant narratives:

$$M_\sigma(a, \{i\}, true) = q \cdot \mathbf{1}[a = g] \cdot f(i, a)$$
$$M_\sigma(a, \{i\}, denial) = q \cdot p_\sigma(a = g) \cdot f(i, a)$$
$$M_\sigma(a, \{i\}, tribal) = p_\sigma(y = G \mid x_i = 1) \cdot f(i, a).$$

The proof proceeds in steps. As a preliminary observation, we note that there must exist $(a, C, S) \in Supp(\sigma)$ such that $a = g$. A formal argument for this appears in the proof of our main result (Theorem 1) below. Intuitively, the trembles of ε-equilibria ensure that the total mobilization generated by the platform $(g, \{1\}, \{0\})$ is $q \cdot f(1, g) > 0$. Therefore, the equilibrium platforms have to generate positive mobilization, which is impossible if policy g is never taken and, hence, the outcome is never G.

STEP 1 (platform carried by true narrative). (i) *If* $\sigma(a, \{i\}, true) > 0$, *then* $a = g$ *and* $i = 1$. (ii) *If* $\sigma(g, \{i\}, S) > 0$, *then* $S = true$.

Proof. Consider an ε-equilibrium σ. Note that $p_\sigma(y = G \mid a = b) = 0$ and $p_\sigma(y = G \mid a = g) = q$. It follows that if $\sigma(a, \{i\}, true) > \varepsilon$ and hence $(a, \{i\}, true)$ maximizes M_σ, then $a = g$ and $i = 1$ because $f(1, g) > f(2, g)$. Now suppose $\sigma(g, \{i\}, S) > \varepsilon$. Since σ has full-support, $p_\sigma(y = G \mid x_{S'}) < q$ whenever $0 \notin S'$. This means that $M_\sigma(g, \{i\}, true) > M_\sigma(g, \{i\}, S')$ for every such S'; hence, $S = true$. We have thus established that claims (i) and (ii) hold for any ε-equilibrium and, hence, in any limit of ε-equilibria. □

Step 1 implies that $(g, \{1\}, \{0\}) \in Supp(\sigma)$, and that if $(a, \{i\}, denial)$ or $(a, \{i\}, tribal)$ are in $Supp(\sigma)$, then $a = b$.

STEP 2 (platforms carried by denial and tribal narratives).
(i) *If* $\sigma(b, \{i\}, denial) > 0$, *then* $i = 2$.
(ii) *If* $\sigma(b, \{i\}, tribal) > 0$, *then* $i = 1$.

Proof. Claim (i) follows immediately from $f(2, b) > f(1, b)$. As to claim (ii), Step 1(i) and $Pr(y = 1|a = b) = 0$ imply that $p_\sigma(y = G \mid x_i = 1) > 0$ only if $i = 1$. Therefore, if $(b, \{i\}, tribal)$ is in $Supp(\sigma)$, then $i = 1$. □

The previous steps pin down the three platforms that can be in $Supp(\sigma)$ for any equilibrium σ, namely $(g, \{1\}, true)$, $(b, \{2\}, denial)$, and $(b, \{1\}, tribal)$. Since they all have distinct narratives, it will be convenient hereafter to *denote each platform by*

its narrative. The total mobilization they generate is

$$\begin{aligned} M_\sigma(true) &= q \cdot f(1,g) \\ M_\sigma(denial) &= q \cdot \sigma(true) \cdot f(2,b) \\ M_\sigma(tribal) &= q \cdot \frac{\sigma(true)}{\sigma(true)+\sigma(tribal)} \cdot f(1,b). \end{aligned} \tag{A.1}$$

STEP 3 (hierarchy of narratives). *In equilibrium,* $\sigma(tribal) > 0$ *only if* $\sigma(denial) > 0$.

Proof. Suppose $\sigma(tribal) > 0 = \sigma(denial)$. Then,

$$\sigma(true) + \sigma(tribal) = 1,$$

so that

$$M_\sigma(tribal) = q \cdot \sigma(true) \cdot f(1,b).$$

But $f(2,b) > f(1,b)$ then implies that $M_\sigma(tribal) < M_\sigma(denial)$, which contradicts $\sigma(tribal) > 0$. □

Steps 1–3 enable us to establish equilibrium existence and uniqueness. Since $\sigma(true) > 0$, every platform in the support of σ generates a total mobilization of $q \cdot f(1,g)$. This requirement reduces the task of deriving σ to solving systems of linear equations under various configurations of f, which determine whether $Supp(\sigma)$ is {true, denial, tribal}, {*true, denial*}, or {*true*}.

Case I: $f(2,b) > f(1,b) > f(1,g) > f(2,g)$. In this case, $M_\sigma(true) < M_\sigma(denial)$ if $\sigma(true) = 1$. Therefore, $\sigma(true) < 1$. It follows from Step 3 that $\sigma(denial) > 0$. Moreover, $\sigma(tribal) > 0$ because otherwise $M_\sigma(tribal) > M_\sigma(true)$. Therefore, σ must satisfy

$$M_\sigma(denial) = M_\sigma(true) = M_\sigma(tribal),$$

which has the unique solution

$$\begin{aligned} \sigma(true) &= \frac{f(1,g)}{f(2,b)} \\ \sigma(denial) &= \frac{f(2,b) - f(1,b)}{f(2,b)} \\ \sigma(tribal) &= \frac{f(1,b) - f(1,g)}{f(2,b)}. \end{aligned}$$

Case II: $f(1,g) \geq f(2,b)$. In this case, $M_\sigma(true) > M_\sigma(denial)$ whenever $\sigma(true) < 1$. It follows that $Supp(\sigma) = \{true\}$. Indeed, when $\sigma(true) = 1$,

$$M_\sigma(true) \geq M_\sigma(denial), M_\sigma(tribal).$$

Thus, $\sigma(true) = 1$ is the unique equilibrium.

Case III: $f(2,b) > f(1,g) \geq f(1,b)$. In this case, $M_\sigma(true) < M_\sigma(denial)$ if $\sigma(true) = 1$. Therefore, $\sigma(true) < 1$. It follows from Step 3 that $\sigma(denial) > 0$. Since $f(1,g) \geq f(1,b)$, then $M_\sigma(tribal) < M_\sigma(true)$ whenever $\sigma(tribal) > 0$. Therefore,

$$\sigma(true) = \frac{f(1,g)}{f(2,b)} \qquad \sigma(denial) = \frac{f(2,b) - f(1,g)}{f(2,b)}$$

is the unique solution of

$$M_\sigma(denial) = M_\sigma(true) \geq M_\sigma(tribal).$$

This completes the proof.

A.2. Proof of Theorem 1

We organize the proof in steps. We will posit the existence of an equilibrium, characterize its properties, and then confirm that we indeed have an equilibrium. Hereafter, let σ be any candidate equilibrium. Note that by definition, $F(N,a) = F(N^a,a)$. We use the two notations interchangeably. For convenience, let

$$d = F(N,b) - F(N,g). \tag{A.2}$$

STEP 1. *There exists* $(a,C,S) \in Supp(\sigma)$ *such that* $a = g$.

Proof. Assume the contrary—that is, $a = b$ for every $(a,C,S) \in Supp(\sigma)$. Then, $p_\sigma(y = G) = 0$. Therefore,

$$M_\sigma(a,C,S) = p_\sigma(y = G \mid x_S(a,C)) = 0,$$

for every $(a,C,S) \in Supp(\sigma)$. By the definition of equilibrium, σ is the limit of a sequence of ε-equilibria for some $\varepsilon \to 0$. Since $\sigma(a,C,S) > 0$, $\sigma_\varepsilon(a,C,S)$ is bounded away from zero, and therefore $M_{\sigma_\varepsilon}(a,C,S) \approx p_{\sigma_\varepsilon}(y = G \mid x_S(a,C)) \approx 0$, for some point along the sequence $\varepsilon \to 0$. In contrast, $M_{\sigma_\varepsilon}(g,N^g,\{0\}) = q \cdot F(N,g)$, which is bounded away from zero and therefore higher than $M_{\sigma_\varepsilon}(a,C,S)$. This contradicts $(g,N^g,\{0\}) \notin Supp(\sigma)$. □

STEP 2. *If* $(g,C,S) \in Supp(\sigma)$, *then* $C = N^g$ *and* $0 \in S$.

Proof. Since $F(N^g, g) > F(C', g)$ for every $C' \subset N^g$, it follows that $C = N^g$ for every $(g, C, S) \in Supp(\sigma)$. Moreover, note that

$$p_\sigma(y = G \mid x_S(g, C)) = q \cdot p_\sigma(x_0 = g \mid x_S(g, C)) \leq q = p_\sigma(y = G \mid x_0 = g).$$

In particular, the inequality is strict if σ has full support, which is the case in ε-equilibrium. Therefore, for every ε-equilibrium $\sigma_\varepsilon(g, N^g, S) \leq \varepsilon$ for all $S \neq \{0\}$. We conclude that $(g, N^g, S) \in Supp(\sigma)$ implies $0 \in S$. □

Comment on the Role of Trembles. The last step establishes part (ii) in the statement of the theorem. Steps 1–2 are the only place in the proof where we use the trembles of ε-equilibria. (The same holds for Step 1 in the proof of Proposition 1.) The trembles ensure that g is implemented with positive probability in equilibrium. (Otherwise, one could sustain a trivial equilibrium in which only b is implemented, using off-path beliefs that g would have equally bad outcomes). From now on, we focus on the $\varepsilon \to 0$ limit itself. Hence, none of the subsequent steps rely on the trembling-hand aspect of our equilibrium concept.

COROLLARY A.1 *Total equilibrium mobilization is equal to*

$$M^* \equiv q \cdot F(N^g, g). \tag{A.3}$$

This follows immediately from Steps 1 and 2. Note that M^* is independent of σ. Denote

$$\alpha = \sigma(g, N^g, \{0\}). \tag{A.4}$$

STEP 3. *If* $x_S(b, C) = x_S(g, N^g)$, *then*

$$p_\sigma(y = G \mid x_S(b, C)) = \frac{q\alpha}{\alpha + \sum_{C', S' \mid x_S(b, C') = x_S(b, C)} \sigma(b, C', S')}. \tag{A.5}$$

Otherwise, $p_\sigma(y = G \mid x_S(b, C)) = 0$.

Proof. Suppose $0 \notin S$. By definition,

$$p_\sigma(y = G \mid x_S(b, C)) = \frac{q \cdot \sum_{C', S' \mid x_S(g, C') = x_S(b, C)} \sigma(g, C', S')}{\sum_{a', C', S' \mid x_S(a', C') = x_S(b, C)} \sigma(a', C', S')}.$$

By Step 2, the numerator can be rewritten as

$$q \cdot \alpha \cdot \mathbf{1}[x_S(b, C) = x_S(g, N^g)],$$

which delivers (A.5). (Note that when $0 \notin S$, $x_S(b, C) = x_S(g, C')$ if and only if $S \cap C = S \cap C'$.) Now suppose $0 \in S$. Then,

$$p_\sigma(y = G \mid x_S(b, C)) = p_\sigma(y = G \mid x_0 = b) = 0. \tag{A.6}$$

□

COROLLARY 3. *For every* $(b, C, S) \in Supp(\sigma)$, $0 \notin S$.

Proof. Suppose $0 \in S$. By (A.6), $M_\sigma(b, C, S) = 0 < M^*$; hence, $(b, C, S) \notin Supp(\sigma)$. □

STEP 4. *If* $F(N, b) \leq F(N, g)$, *then* $\alpha = 1$. *If* $F(N, b) > F(N, g)$, *then*

$$\alpha \leq \frac{F(N, g)}{F(N, b)}.$$

Proof. Suppose $F(N, b) \leq F(N, g)$, but $\alpha < 1$. Then, there exists $(b, C, S) \in Supp(\sigma)$, such that the denominator of (A.5) is greater than α and, hence, $p_\sigma(y = G \mid x_S(b, C)) < q$. It follows that

$$M_\sigma(b, C, S) = p_\sigma(y = G \mid x_S(b, C)) \cdot F(C, b) < q \cdot F(N, b) \leq q \cdot F(N, g) = M^*,$$

which is a contradiction. Thus, in this case $\alpha = 1$. Suppose $F(N, b) > F(N, g)$. If $\alpha = 1$, then

$$M_\sigma(b, N^b, \varnothing) = p_\sigma(y = G)F(N, b) = qF(N, b) > M^*,$$

which is a contradiction. Thus, in this case $\alpha < 1$. Recall that the denial narrative $S = \varnothing$ is feasible. Furthermore, we must have $M_\sigma(b, N^b, \varnothing) \leq M^*$ in any equilibrium. Since $p_\sigma(y = G) = q\alpha$, it follows that $q\alpha \cdot F(N, b) \leq q \cdot F(N, g)$. This implies the upper bound on α when $\alpha < 1$. □

Steps 1 and 4 establish part (i) in the statement of the theorem. The next step proves part (iii).

STEP 5. *If* $(b, C, S) \in Supp(\sigma)$, *then* $L(S) \subseteq N \setminus N^g$ *and* $C = N^b \setminus L(S)$.

Proof. We first show that $N^g \cap N^b \subseteq C$ for every $(a, C, S) \in Supp(\sigma)$, and then use this observation to establish the claim. Assume there is a platform $(a, C, S) \in Supp(\sigma)$ such that $j \notin C$ for some $j \in (N^g \cap N^b)$. By Step 2, $a = b$. There are two cases to consider: *Case 1*: $j \notin S$. Then, $p_\sigma(y = G \mid x_S(b, C \cup \{j\})) = p_\sigma(y = G \mid x_S(b, C))$. But since $F(C \cup \{j\}, b) > F(C, b)$, it follows that $M_\sigma(b, C \cup \{j\}, S) > M_\sigma(b, C, S)$, a contradiction. *Case 2*: $j \in S$. Since $x_j(a, C) = 0$ and every platform with $a = g$ includes j in its coalition, we have that $p_\sigma(y = G \mid x_S(a, C)) = 0$. But then $(b, C, S) \notin Supp(\sigma)$, a contradiction. We have thus shown that the Center is always in every ruling coalition. Consider some platform $(b, C, S) \in Supp(\sigma)$. By assumption, no $j \in N \setminus N^b$ is in C. From the argument above, $(N^g \cap N^b) \subseteq C$. In addition, $0 \notin S$ and $S \cap (N \setminus N^b) = \varnothing$ as otherwise, $p_\sigma(y = G \mid x_S(b, C)) = 0$. It follows that $S \setminus (N^g \cap N^b) \subseteq N \setminus N^g$ (this includes the case where $S \setminus (N^g \cap N^b) = \varnothing$.) It remains to show that $C = N^b \setminus L(S)$. First, suppose there is $j \in L(S)$ such that $j \in C$. Then, $x_j(b, C) = 1$ and, hence, $p_\sigma(y = G \mid x_S(b, C)) = 0$ (since

j is not in any coalition that is part of a platform with $a = g$), a contradiction. Second, suppose there is $j \in N \setminus N^g$ such that $j \notin S$ and $j \notin C$. Then, since $p_\sigma(y = G \mid x_S(b, C \cup \{j\})) = p_\sigma(y = G \mid x_S(b, C))$ and $F(C \cup \{j\}, b) > F(C, b)$, it follows that $M_\sigma(b, C \cup \{j\}, S) > M_\sigma(b, C, S)$, a contradiction. □

The rest of the proof establishes uniqueness of the equilibrium distribution over (a, C), and provides an algorithm for computing it (which will be put to use in subsequent results).

The last step implies that the equilibrium probability of a pair (b, C) is entirely pinned down by C. In particular, any platform $(b, C, S) \in Supp(\sigma)$ satisfies $C = N^b \setminus L(S)$. We use this observation to introduce the following notation, which we will use for the remainder of the proof. Let $\mathcal{S}$ denote the domain of feasible tribal narratives, and let $\mathcal{T} \equiv \{L(S) \mid S \in \mathcal{S}\}$. For every $T \in \mathcal{T}$, define

$$\bar{\sigma}(T) \equiv \sum_{C,S|L(S)=T} \sigma(b, C, S). \tag{A.7}$$

STEP 6. *There is an equilibrium σ that induces the distribution $(\alpha, \bar{\sigma})$ if and only if, for all $T \in \mathcal{T}$ that satisfy $T \subseteq N \setminus N^g$,*

$$\alpha \cdot \frac{d - F(T, b)}{F(N, g)} \leq \sum_{T' \in \mathcal{T} | T' \supseteq T} \bar{\sigma}(T'), \tag{A.8}$$

with equality if $\bar{\sigma}(T) > 0$. (Recall that d is defined by (A.2).)

Proof. By Definition 2, σ is an equilibrium if and only if $M_\sigma(b, C, S) \leq M^*$ for all (b, C, S), with equality if $\sigma(b, C, S) > 0$. By Corollary 2 and Step 3, this inequality can be written as follows:

$$\frac{\alpha \cdot F(C, b)}{\alpha + \sum_{C', S' | x_S(b,C') = x_S(b,C)} \sigma(b, C', S')} \leq F(N, g). \tag{A.9}$$

By Step 5, $C = N^b \setminus L(S)$. Therefore, the above inequality reduces to a linear inequality in σ:

$$\alpha \cdot \frac{d - F(L(S), b)}{F(N, g)} \leq \sum_{C', S' | x_S(b,C') = x_S(b,C)} \sigma(b, C', S'). \tag{A.10}$$

Again, by Step 5, if $\sigma(b, C', S') > 0$, then $C' = N^b \setminus L(S')$, such that $x_S(b, C') = x_S(b, C)$ if and only if $L(S') \supseteq L(S)$. This means that we can replace the R.H.S. of the last inequality with the R.H.S. of (A.8). □

Inequalities (A.8) enable us to construct the following algorithm that associates with every equilibrium σ a unique distribution over $\bar{\sigma}(T)$ for every $T \in \mathcal{T}$ satisfying $T \in N \setminus N^g$.

The Algorithm. Let

$$\overline{\mathcal{T}} = \{T \in \mathcal{T} \mid T \subseteq N \setminus N^g \text{ and } F(T,b) < d\}.$$

Define

$$\overline{\mathcal{T}}_1 = \{T \in \overline{\mathcal{T}} \mid \text{there is no } T' \in \overline{\mathcal{T}} \text{ such that } T \subset T'\}.$$

Now, for every $k > 1$, define $\overline{\mathcal{T}}_k$ recursively as follows:

$$\overline{\mathcal{T}}_k = \{T \in \overline{\mathcal{T}} \mid \text{there is no } T' \in \overline{\mathcal{T}} \setminus \cup_{j<k} \overline{\mathcal{T}}_j \text{ such that } T \subset T'\}.$$

Since $\overline{\mathcal{T}}$ is finite, in this way we obtain a finite sequence $\{\overline{\mathcal{T}}_k\}_{k=1}^K$. This sequence identifies all the "exclusionary" components of feasible narratives (i.e., those that scapegoat groups in $N \setminus N^g$) that can accompany platforms with a policy of b.

The algorithm starts from the "top layer" of $\overline{\mathcal{T}}$ (i.e., $\overline{\mathcal{T}}_1$) and then proceeds to the other layers in order. For every $T \in \overline{\mathcal{T}}_1$, (A.8) can be written as

$$\bar{\sigma}(T) \geq \alpha \cdot \frac{d - F(T,b)}{F(N,g)}.$$

By the definition of $\overline{\mathcal{T}}$, the R.H.S. is strictly positive for every $T \in \overline{\mathcal{T}}_1$, which implies that T is in the equilibrium support and therefore the inequality must hold with equality. This pins down $\bar{\sigma}(T)$.

For every $T \in \overline{\mathcal{T}}$, denote $\mathcal{H}(T) \equiv \{T' \in \overline{\mathcal{T}} \mid T \subset T'\}$. By definition, if $T \in \overline{\mathcal{T}}_k$, then $\mathcal{H}(T) \subseteq \cup_{j<k} \overline{\mathcal{T}}_j$. We proceed by induction. Suppose that for all $j < k$ and every $T \in \overline{\mathcal{T}}_j$, there exists $w(T) \geq 0$ such that

$$\bar{\sigma}(T) = \alpha w(T).$$

For $T \in \overline{\mathcal{T}}_1$, we have already established that $w(T) = (d - F(T,b))/F(N,g)$. For every $T \in \overline{\mathcal{T}}_k$, (A.8) becomes

$$\bar{\sigma}(T) = \max\left\{0\,,\ \alpha \cdot \frac{d - F(T,b)}{F(N,g)} - \alpha \sum_{T' \in \mathcal{H}(T)} w(T')\right\}, \tag{A.11}$$

where $w(T')$ is well-defined for all $T' \in \mathcal{H}(T)$, by the inductive step. This confirms that $\bar{\sigma}(T) = \alpha w(T)$, where

$$w(T) = \max\left\{0\,,\ \frac{d - F(T,b)}{F(N,g)} - \sum_{T' \in \mathcal{H}(T)} w(T')\right\}, \tag{A.12}$$

completing the inductive argument, and thus the definition of the algorithm for computing $\bar{\sigma}(T)$.

STEP 7. The algorithm establishes existence of an equilibrium σ and uniqueness of the induced distribution $(\alpha, \bar{\sigma})$.

Proof. Since $(\alpha, \bar{\sigma})$ must define a probability distribution, we must have

$$\alpha + \sum_{T \in \overline{\mathcal{T}}} \bar{\sigma}(T) = 1.$$

Moreover, the algorithm produced unique expressions for each $\bar{\sigma}(T)$ that depend multiplicatively on α (see (A.11) and (A.12)). This pins down the value of α,

$$\alpha = \frac{1}{1 + \sum_{T \in \overline{\mathcal{T}}} w(T)}.$$

Thus, we have pinned down $(\alpha, \bar{\sigma})$. Since this pair satisfies all the inequalities (A.8), it implies that the following distribution over platforms is an equilibrium: $\alpha = \sigma(g, N^g, 0)$ and $\bar{\sigma}(T) = \sigma(b, N^b \setminus T, T)$ for every $T \in \mathcal{T}$ such that $T \in N \setminus N^g$. □

A.3. Proof of Proposition 2

This result is a corollary of Step 6 in the proof of Theorem 1. Suppose $0 < F(T) < F(N, b) - F(N, g)$ for some $T \subseteq N \setminus N^g$. Then, the L.H.S. of (A.8) is strictly positive. Therefore, we must have $\bar{\sigma}(T') > 0$ for some such $T' \supseteq T$. Conversely, suppose $F(T) \geq F(N, b) - F(N, g)$ for all $T \subseteq N \setminus N^g$. In this case, the L.H.S of (A.8) is non-positive for every such T. By Step 6, this implies $\bar{\sigma}(T) = 0$ for every such T.

A.4. Proof of Theorem 2

Let $\mathcal{S}^*$ be the collection of coarse subcategories of the Left—that is, a feasible tribal narrative $S \subset N \setminus N^g$ is in $\mathcal{S}^*$ if there is no $S' \in \mathcal{S}$ such that $S \subset S' \subset N \setminus N^g$. Let

$$\mathcal{S}^{\neg *} = \{S \in \mathcal{S} \mid S \subset N \setminus N^g \text{ and } S \notin \mathcal{S}^*\}.$$

For every $S \in \mathcal{S}$, let $B(S) = N \setminus (N^g \cup S)$—that is, $B(S)$ is the set of Left groups that do not belong to S. Finally, recall that we are focusing on essential equilibria.

We use the notation $\bar{\sigma}$ as in the proof of Theorem 1. By (A.8) and (5),

$$\bar{\sigma}(N \setminus N^g) = \alpha \cdot \frac{F(N^g, b) - F(N, g)}{F(N, g)} > 0. \tag{A.13}$$

Also, for $S \in \mathcal{S}^*$, we have

$$\bar{\sigma}(S) = \alpha \cdot \frac{d - F(S, b)}{F(N, g)} - \bar{\sigma}(N \setminus N^g) = \alpha \cdot \frac{F(N \setminus N^g, b) - F(S, b)}{F(N, g)} > 0. \tag{A.14}$$

These expressions establish that the Left and its coarse sub-categories are employed with positive probability as tribal narratives in every essential equilibrium. The following lemma establishes that under property (ii), these are the only non-empty tribal narratives that are employed.

LEMMA A.1. *If property* (ii) *holds, then* $\bar{\sigma}(S) = 0$ *for every non-empty* $S \in \mathcal{S}^{\neg *}$.

Proof. Assume the contrary—that is, property (ii) holds and yet there is $S \in \mathcal{S}^{\neg *}$ such that $\bar{\sigma}(S) > 0$. Select S such that there is no $S' \in \mathcal{S}^{\neg *}$ for which $S \subset S'$ and $\bar{\sigma}(S') > 0$. We have

$$\begin{aligned}\bar{\sigma}(S) &\geq \alpha \cdot \frac{d - F(S,b)}{F(N,g)} - \bar{\sigma}(N \setminus N^g) - \sum_{S' \in \mathcal{S}^* | S \subset S'} \bar{\sigma}(S') \\ &= \alpha \cdot \left(\frac{d - F(S,b)}{F(N,g)} - \frac{F(N^g,b) - F(N,g)}{F(N,g)} \right. \\ &\quad \left. - \sum_{S' \in \mathcal{S}^* | S \subset S'} \frac{F(N \setminus N^g,b) - F(S',b)}{F(N,g)} \right) \\ &= \frac{\alpha}{F(N,g)} \cdot \left(F(B(S),b) - \sum_{S' \in \mathcal{S}^* | S \subset S'} F(B(S'),b) \right), \end{aligned} \tag{A.15}$$

where the inequality follows from (A.8), and the subsequent equations result from using (A.13) and (A.14). If $\mathcal{S}$ satisfies property (ii), then

$$B(S) \subseteq \bigcup_{S' \in \mathcal{S}^* | S \subset S'} B(S'),$$

which implies that the difference in (A.15) is weakly negative. Hence, $\bar{\sigma}(S) = 0$, a contradiction. □

Part I ("if"). Suppose properties (i) and (ii) hold. Taken together, Lemma A.1 and equations (A.13) and (A.14) state that $N \setminus N^g$ and all $S \in \mathcal{S}^*$ are in $Supp(\bar{\sigma})$, and that $Supp(\bar{\sigma})$ includes no other non-empty $S \subset N \setminus N^g$.

If $\bar{\sigma}(\varnothing) > 0$, then (A.8) becomes

$$\alpha \cdot \frac{d - F(\varnothing,b)}{F(N,g)} = \sum_{S \supseteq \varnothing} \bar{\sigma}(S) = 1 - \alpha,$$

which implies $\alpha = F(N,g)/F(N,b)$.

Now suppose $\bar{\sigma}(\varnothing) = 0$. Then,

$$1 = \alpha + \bar{\sigma}(N \setminus N^g) + \sum_{S' \in \mathcal{S}^*} \bar{\sigma}(S'). \tag{A.16}$$

By the same calculation as in (A.15),

$$\bar{\sigma}(\varnothing) \geq \frac{\alpha}{F(N,g)} \cdot \left(F(N \setminus N^g,b) - \sum_{S' \in \mathcal{S}^*} F(B(S'),b) \right).$$

Since $\mathcal{S}$ satisfies property (i), $B(S') \cap B(S'') = \varnothing$ for every $S', S'' \in \mathcal{S}^*$. This implies that the R.H.S. of the last inequality is non-negative. And since $\bar{\sigma}(\varnothing) = 0$, the R.H.S. must be exactly zero. Using this observation and plugging (A.13) and (A.14) into (A.16), we obtain

$$1 = \alpha \left(\frac{F(N^g, b)}{F(N, g)} + \frac{F(N \setminus N^g, b)}{F(N, g)} \right) = \alpha \frac{F(N, b)}{F(N, g)},$$

which again implies $\alpha = F(N, g)/F(N, b)$. Note that we reach this conclusion for any f and, hence, for f such that $F(N^g \cap N^b, b) > F(N, g)$.

Part II ("only if"). Suppose property (i) does not hold. Equations (A.13) and (A.14) continue to hold. In particular, $N \setminus N^g$ and every $S \in \mathcal{S}^*$ are in $Supp(\bar{\sigma})$. Note that

$$1 \geq \alpha + \bar{\sigma}(N \setminus N^g) + \sum_{S \in \mathcal{S}^*} \bar{\sigma}(S).$$

Plugging (A.13) and (A.14) in the R.H.S. yields

$$1 \geq \frac{\alpha}{F(N, g)} \left(F(N^g, b) + \sum_{S \in \mathcal{S}^*} F(B(S), b) \right).$$

Therefore, $\alpha < F(N, g)/F(N, b)$ if

$$F(N^g, b) + \sum_{S \in \mathcal{S}^*} F(B(S), b) > F(N, b). \tag{A.17}$$

We claim that there exist values of $F(N \setminus N^g, b)$ for which this happens, while holding $F(N, g)$ and $F(N^g \cap N^b, b)$ fixed. Since property (i) fails, there exist $S, S' \in \mathcal{S}^*$ such that $B(S) \cap B(S') \neq \varnothing$. Thus, every i in this intersection is counted more than once on the L.H.S. of (A.17). We can then choose f such that, for any $i \in B(S) \cap B(S')$,

$$f(i, b) > F(N \setminus N^g, b) - F(B(\mathcal{S}^*), b) = F\big(N \setminus (N^g \cup B(\mathcal{S}^*)), b\big),$$

where $B(\mathcal{S}^*) \equiv \cup_{S \in \mathcal{S}^*} B(S)$.

Now, suppose property (i) holds but property (ii) fails. This failure implies that there exists a non-empty $S \in \mathcal{S}^{\neg *}$ such that[14]

$$B(S) \supset \bigcup_{S' \in \mathcal{S}^* | S \subset S'} B(S'). \tag{A.18}$$

Moreover, we claim that there exists a non-empty $S \in \mathcal{S}^{\neg *}$ that satisfies (A.18) and $\bar{\sigma}(S) > 0$. Suppose not. From Part I of this proof, we know that $\bar{\sigma}(S') = 0$ if $S' \in \mathcal{S}^{\neg *}$ satisfies property (ii). Therefore, for any non-empty $S \in \mathcal{S}^{\neg *}$ that satisfies (A.18), we

14. Note that if there is no non-empty $S \in \mathcal{S}^{\neg *}$, then property (ii) cannot fail. In this case, the proof is complete.

can write

$$\begin{aligned}\bar{\sigma}(S) &\geq \alpha \cdot \frac{d - F(S,b)}{F(N,g)} - \bar{\sigma}(N \setminus N^g) - \sum_{S' \in \mathcal{S}^* : S \subset S'} \bar{\sigma}(S') \\ &= \alpha \cdot \left(\frac{d - F(S,b)}{F(N,g)} - \frac{F(N^g,b) - F(N,g)}{F(N,g)} - \sum_{S' \in \mathcal{S}^* | S \subset S'} \frac{F(B(S'),b)}{F(N,g)} \right) \\ &= \alpha \cdot \left(\frac{F(B(S),b)}{F(N,g)} - \sum_{S' \in \mathcal{S}^* | S \subset S'} \frac{F(B(S'),b)}{F(N,g)} \right) > 0,\end{aligned}$$

where the strict inequality follows using (A.18) and property (i) (which means that $B(S') \cap B(S'') = \varnothing$ for all distinct $S', S'' \in \mathcal{S}^*$ such that $S \subset S', S''$). This contradicts the premise that $\bar{\sigma}(S) = 0$, proving our claim.

Now take any $S' \in \mathcal{S}^{\neg *}$ such that $\bar{\sigma}(S') > 0$. Note that

$$\begin{aligned}1 &\geq \alpha + \bar{\sigma}(N \setminus N^g) + \sum_{S \in \mathcal{S}^*} \bar{\sigma}(S) + \bar{\sigma}(S') \\ &= \frac{\alpha}{F(N,g)} \left(F(N^g,b) + \sum_{S \in \mathcal{S}^*} F(B(S),b) + F(B(S'),b) \right).\end{aligned}$$

Therefore, $\alpha < F(N,g)/F(N,b)$ if

$$F(N^g,b) + \sum_{S \in \mathcal{S}^*} F(B(S),b) + F(B(S'),b) > F(N,b). \tag{A.19}$$

We again claim that there exist values of $F(N \setminus N^g, b)$ for which this inequality holds while keeping $F(N,g)$ and $F(N^g \cap N^b, b)$ fixed. The reason is that since S' satisfies (A.18), there exists

$$i \in B(S') \cap \bigcup_{S \in \mathcal{S}^* | S' \subset S} B(S),$$

that is counted more than once on the L.H.S. of (A.19). Therefore, we can choose such i and set $f(i,b)$ such that

$$f(i,b) > F\big(N \setminus \big(N^g \cup B(S^*)\big), b\big).$$

This completes the proof.

A.5. Proof of Proposition 3

Let σ be the unique essential equilibrium. Since $F(N,b) > F(N^g,b) > F(N,g)$, Theorem 1 implies that $\sigma(g, N^g, \{0\}) = \alpha \in (0,1)$. Let us now activate the algorithm described in the proof of Proposition 1. The restriction to essential equilibria allows us to identify any equilibrium platform with its narrative. Therefore, we will use

the abbreviated notation $\bar{\sigma}(S) = \sigma(b, C, S)$. Also, for every $S \subset N \setminus N^g$, denote $S^c = (N \setminus N^g) \setminus S$.

As in the proof of Theorem 2, $\bar{\sigma}(N \setminus N^g)$ is given by (A.13). Now consider the largest feasible tribal narratives $S \subset N \setminus N^g$. By definition, these take the form

$$S = (N \setminus N^g) \cap \{i \in N \mid i_k = w\}, \tag{A.20}$$

where $k \in \{1, \ldots, m\}$ and $w \in \{0, 1\}$. Denote this set of $2m$ narratives by $\mathcal{S}^*$. By definition, $S \not\subseteq S'$ for any $S' \neq S$ such that $S' \subset N \setminus N^g$. Therefore, if $\bar{\sigma}(S) = 0$ for some $S \in \mathcal{S}^*$ then the following inequality must hold:

$$\alpha \cdot \frac{F(N^g \cup S^c, b) - F(N, g)}{F(N, g)} \leq \bar{\sigma}(N \setminus N^g),$$

which is a contradiction since $F(N^g \cup S^c, b) > F(N^g, b)$. It follows that for every $S \in \mathcal{S}^*$,

$$\bar{\sigma}(S) = \alpha \cdot \frac{F(S^c, b)}{F(N, g)} > 0. \tag{A.21}$$

The support of $\bar{\sigma}$ contains no other narratives. To see why, recall that in Section 5.2, we explained why the multi-attribute model satisfies property (ii). Therefore, applying Lemma A.1, we conclude that the support of $\bar{\sigma}$ consists of the true narrative (whose equilibrium probability is α), $N \setminus N^g$ and all the narratives in $\mathcal{S}^*$. By (A.13) and (A.21),

$$\alpha + \alpha \cdot \frac{F(N^g, b) - F(N, g)}{F(N, g)} + \alpha \cdot \frac{1}{F(N, g)} \sum_{S \in \mathcal{S}^*} F(S^c, b) = 1. \tag{A.22}$$

By definition,

$$F(S, b) + F(S^c, b) = F(N \setminus N^g, b)$$

for every $S \in \mathcal{S}^*$. Therefore,

$$\sum_{S \in \mathcal{S}^*} F(S^c, b) = m \cdot F(N \setminus N^g, b),$$

so that (A.22) implies (6).

A.6. Proof of Proposition 4

As explained in Section 5.3, every feasible $S \subseteq N \setminus N^g$ is employed as an exclusionary tribal narrative in the essential equilibrium. We will take this feature for granted and use the algorithm in the proof of Theorem 1 to derive the equilibrium probabilities of all such narratives.

It will be convenient to translate the hierarchical multi-attribute model into a system Π of nested partitions of the set $N \setminus N^g$. Let $\pi_0 = \{N \setminus N^g\} = \{\{i \in N \mid i_k = 1 \text{ for all } k > m\}\}$. For every $\ell = 1, \ldots, D$, let π_ℓ consist of all sets of the form $S \cap \{i \in N \mid i_{m-\ell+1} = v\}$, where $S \in \pi_{\ell-1}$ and $v \in \{0, 1\}$. Thus, for instance, π_1

consists of the two cells $N \setminus N^g \cap \{i \in N \mid i_m = 1\}$ and $N \setminus N^g \cap \{i \in N \mid i_m = 0\}$.

We make use of the same abbreviated notation $\bar{\sigma}$ as in the proof of Proposition 3. As in that case,

$$\bar{\sigma}(N \setminus N^g) = \alpha \cdot \frac{F(N^g, b) - F(N, g)}{F(N, g)}.$$

This characterizes the equilibrium probability of the single cell that comprises π_0. Now consider $\ell > 1$. Given $S_\ell \in \pi_\ell$, the collection of sets $\mathcal{H}(S_\ell) = \{S' \in \overline{\mathcal{S}} \mid S_\ell \subset S'\}$ in the algorithm described in the proof of Theorem 1 takes the form of a chain $\{S_j\}_{j=1}^{\ell-1}$ that satisfies $S_j \in \pi_j$ and $S_{j+1} \subset S_j$ for all $j < \ell$. For $S_1 \in \pi_1$, we must have

$$\bar{\sigma}(S_1) = \frac{\alpha(d - F(S_1, b)) - \alpha(d - F(S_1, b))}{F(N, g)} = \alpha \frac{F(S_1 \setminus S_2, b)}{F(N, g)}.$$

Thus, the coefficient, $w(S_2)$ in the proof of Theorem 1 is, takes the form $F(S_1 \setminus S_2, b)/F(N, g)$. By induction,

$$\bar{\sigma}(S_\ell) = \alpha \frac{F(S_{\ell-1} \setminus S_\ell, b)}{F(N, g)}, \tag{A.23}$$

for every $S_\ell \in \pi_\ell$, $\ell = 1, \ldots, D$. This completes the characterization of the $\bar{\sigma}(S)$ for every cell S in one of the nested partitions in Π.

Before the final step of the proof, it also needs to be shown that $\bar{\sigma}(\varnothing) = 0$. The calculation that establishes this is straightforward but somewhat tedious, and we omit it for brevity. The intuition is that while every cell in one of the nested partitions is contained by a relatively small number of other cells, $\varnothing$ is contained by all of these cells. As a result, the R.H.S. of (A.8) is too large for this inequality to be binding for $S = \varnothing$, which means that $\bar{\sigma}(\varnothing) = 0$.

It remains to calculate α. For every $S_\ell \in \pi_\ell$, let $S_{\ell-1}$ be again the antecedent of S_ℓ in the chain $\{S_j\}_{j=1}^{\ell-1}$ that we used above. For every $S \in \pi_\ell$, let $P(S)$ be the unique cell $S' \in \pi_{\ell-1}$ such that $S \subset S'$. Given this, and plugging (A.23), we have

$$\begin{aligned}
1 &= \alpha + \sum_{S \subseteq N \setminus N^g} \bar{\sigma}(S) \\
&= \frac{\alpha}{F(N, g)} \left\{ F(N, g) + d - F(N \setminus N^g, b) + \sum_{\ell=1}^{D} \sum_{S \in \pi_\ell} F(P(S) \setminus S, b) \right\} \\
&= \frac{\alpha}{F(N, g)} \left\{ F(N^g, b) + \sum_{\ell=1}^{D} \sum_{S \in \pi_\ell} F(P(S) \setminus S, b) \right\}.
\end{aligned}$$

To further simplify this expression, we now use the assumption that each cell in $\pi_{\ell-1}$ has exactly two subsets in π_ℓ. Using this, we can rewrite the last condition as

$$1 = \frac{\alpha}{F(N, g)} \{F(N^g, b) + D \cdot F(N \setminus N^g, b)\},$$

which implies (8).

A.7. Proof of Proposition 6

In this proof, we denote platforms by z whenever convenient to simplify notation. For every t, let $\bar{z}_t = (\bar{a}_t, w\bar{C}_t, \bar{S}_t) \in \arg\max_z M_{\sigma_t}(z)$ be the dominant platform at period t and let $\overline{M}_{\sigma_t} = M_{\sigma_t}(\bar{z}_t)$ be the payoff it generates. Note that if there exists T such that $\bar{z}_t \neq (a, C, S)$ for all $t \geq T$, then $\sigma_t(a, C, S) \to 0$ as $t \to \infty$. Recall that $M^* = q \cdot F(N^g, g) > 0$. The proof is organized in four lemmas.

LEMMA A.2. *If $\bar{z}_t = (g, C, S)$, then $C = N^g$ and $M_{\sigma_t}(g, C, S) = M^*$.*

Proof. First, note that $\overline{M}_{\sigma_t} \geq M^*$ for every t. Indeed, since σ_1 has full support, $\sigma_t(g, N^g, \{0\}) > 0$ for every finite t; therefore, $\overline{M}_{\sigma_t} \geq M_{\sigma_t}(g, N^g, \{0\}) = M^*$ for every t. To prove the first implication in the lemma, note that for every platform (g, C, S) such that $C \subset N^g$, $M_{\sigma_t}(g, C, S) < M_{\sigma_t}(g, N^g, \{0\})$ because $Pr_{\sigma_t}(y = G \mid x_S(g, C)) \leq q$ and $F(C, g) < F(N^g, g)$. This also implies that $M_{\sigma_t}(g, N^g, S) \leq M^*$ for all S and hence the last equality in the lemma. □

LEMMA A.3. $\liminf_{t\to\infty} \overline{M}_{\sigma_t} = M^*$.

Proof. Since, as noted, $\overline{M}_{\sigma_t} \geq M^*$ for every t, we have

$$\liminf_{t\to\infty} \overline{M}_{\sigma_t} \geq M^*.$$

Suppose there exists t such that $\bar{z}_{t'} = (b, \bar{C}_{t'}, \bar{S}_{t'})$ for all $t' \geq t$. This implies that $Pr_{\sigma_t}(y = G \mid x_{\bar{S}_t}(\bar{a}_t, \bar{C}_t)) \to 0$, which is inconsistent with $\liminf_{t\to\infty} \overline{M}_{\sigma_t} > 0$. Therefore, for all t, there exists $t' > t$ such that $\bar{z}_{t'} = (g, N^g, S)$ for some S. This property in turn implies the equality in the lemma. To see why, note that, if $\overline{M}_{\sigma_t} > M^*$, then $\bar{z}_t = (b, C, S)$ for some C and S, because $M_{\sigma_t}(g, C', S') \leq M^*$ for all C' and S'. Now suppose $\liminf_{t\to\infty} \overline{M}_{\sigma_t} > M^*$. Then, there must exist T such that for all $t \geq T$, $\bar{z}_t$ involves policy $a = b$, which is a contradiction. □

LEMMA A.4. $\limsup_{t\to\infty} \overline{M}_{\sigma_t} \leq M^*$.

Proof. Recall that

$$Pr_{\sigma_t}(y = G \mid x_S(a, C)) = q \cdot \frac{\sum_{C', S' \mid x_S(g, C') = x_S(a, C)} \sigma_t(g, C', S')}{\sum_{a', C', S' \mid x_S(a', C') = x_S(a, C)} \sigma_t(a', C', S').}$$

To prove this lemma, we first claim four properties of the process σ_t.

CLAIM A.1. If $\bar{z}_t = (g, N^g, \hat{S})$ and $x_S(N^g, g) = x_S(b, C)$, then

$$Pr_{\sigma_{t+1}}(y = G \mid x_S(b, C)) > Pr_{\sigma_t}(y = G \mid x_S(b, C)).$$

Proof. Given $\bar{z}_t = (g, N^g, \hat{S})$, for every (b, C, S) such that $x_S(g, N^g) = x_S(b, C)$,

$$Pr_{\sigma_{t+1}}(y = G \mid x_S(b,C)) = q\frac{\frac{1}{t+1} + \frac{t}{t+1}\sum_{C',S' \mid x_S(g,C') = x_S(b,C)} \sigma_t(g, C', S')}{\frac{1}{t+1} + \frac{t}{t+1}\sum_{a',C',S' \mid x_S(a',C') = x_S(b,C)} \sigma_t(a', C', S')}$$

$$= q\frac{\frac{1}{t} + \sum_{C',S' \mid x_S(g,C') = x_S(b,C)} \sigma_t(g, C', S')}{\frac{1}{t} + \sum_{a',C',S' \mid x_S(a',C') = x_S(b,C)} \sigma_t(a', C', S')}$$

$$> q\frac{\sum_{C',S' \mid x_S(g,C') = x_S(b,C)} \sigma_t(g, C', S')}{\sum_{a',C',S' \mid x_S(a',C') = x_S(b,C)} \sigma_t(a', C', S')} = Pr_{\sigma_t}(y = G \mid x_S(b,C)).$$

□

CLAIM A.2. *If $\bar{z}_t = (b, \hat{C}, \hat{S})$, then for every (b, C, S),*

$$Pr_{\sigma_{t+1}}(y = G \mid x_S(b,C)) \leq Pr_{\sigma_t}(y = G \mid x_S(b,C)),$$

with strict inequality if and only if $x_S(b, \hat{C}) = x_S(b, C)$.

Proof. If $\bar{z}_t = (b, \hat{C}, \hat{S})$ and $x_S(b, \hat{C}) \neq x_S(b, C)$, then by definition, $Pr_{\sigma_{t+1}}(y = G \mid x_S(b,C)) = Pr_{\sigma_t}(y = G \mid x_S(b,C))$. Now suppose that $\bar{z}_t = (b, \hat{C}, \hat{S})$ and $x_S(b, \hat{C}) = x_S(b, C)$. Then,

$$Pr_{\sigma_{t+1}}(y = G \mid x_S(b,C)) = q\frac{\frac{t}{t+1}\sum_{C',S' \mid x_S(g,C') = x_S(b,C)} \sigma_t(g, C', S')}{\frac{1}{t+1} + \frac{t}{t+1}\sum_{a',C',S' \mid x_S(a',C') = x_S(b,C)} \sigma_t(a', C', S')}$$

$$= q\frac{\sum_{C',S' \mid x_S(g,C') = x_S(b,C)} \sigma_t(g, C', S')}{\frac{1}{t} + \sum_{a',C',S' \mid x_S(a',C') = x_S(b,C)} \sigma_t(a', C', S')}$$

$$< q\frac{\sum_{C',S' \mid x_S(g,C') = x_S(b,C)} \sigma_t(g, C', S')}{\sum_{a',C',S' \mid x_S(a',C') = x_S(b,C)} \sigma_t(a', C', S')} = Pr_{\sigma_t}(y = G \mid x_S(b,C)).$$

□

CLAIM A.3. *If (b, C, S) is such that $x_S(b, C) \neq x_S(g, N^g)$, then $\sigma_t(b, C, S) \to 0$ as $t \to \infty$.*

Proof. Suppose $\sigma_t(b, C, S) \not\to 0$. Then, there exists a subsequence such that $\sigma_t(b, C, S) \to \hat{\sigma} > 0$, which implies that the denominator of $Pr_{\sigma_t}(y = G | x_S(b, C))$ converges to a strictly positive number along the subsequence. However, the numerator of $Pr_{\sigma_t}(y = G | x_S(b, C))$ converges to zero by Lemma A.2, because $\sigma_t(g, C', S') \to 0$ if $x_S(g, C') = x_S(b, C)$ and hence C'^g. Therefore, $M_{\sigma_t}(b, C, S) \to 0$ along the subsequence, which contradicts $\sigma_t(b, C, S) \to \hat{\sigma} > 0$. □

CLAIM A.4. *If* (b, C, S) *is such that* $x_S(b, C) = x_S(N^g, g)$, *then*

$$\liminf_{t\to\infty} \sum_{C',S'|x_S(g,C')=x_S(b,C)} \sigma_t(g, C', S') = \liminf_{t\to\infty} \sum_{S'} \sigma_t(g, N^g, S') \equiv \underline{\sigma} > 0.$$

Proof. The first equality follows because $\sigma_t(g, C', S') \to 0$ if C'^g by Lemma A.2 and because $x_S(b, C) = x_S(g, N^g)$. The last inequality is strict because, if $\underline{\sigma} = 0$, there exists a subsequence such that $\sum_{C',S'} \sigma_t(g, C', S') \to 0$ and, hence, $\sigma_t(b, C, S) \to \hat{\sigma} > 0$ for some (b, C, S) such that $x_S(b, C) = x_S(g, N^g)$. However, in this case, there exists T such that for all $t \geq T$ in this subsequence the numerator of $Pr_{\sigma_t}(y = G \mid x_S(b, C))$ becomes arbitrarily small and hence $M_{\sigma_t}(b, C, S) < M^*$, which is inconsistent with $\hat{\sigma} > 0$. □

To complete the proof, suppose $\limsup_{t\to\infty} \overline{M}_{\sigma_t} = \overline{M} > M^*$. Let

$$\overline{P} = \left\{ (b, C, S) \mid \limsup_{t\to\infty} M_{\sigma_t}(b, C, S) = \overline{M} \right\},$$

which must be non-empty because the set of platforms is finite. Note that $(b, C, S) \in \overline{P}$ only if $x_S(b, C) = x_S(g, N^g)$. By finiteness of $\overline{P}$, there exists a common subsequence, T, and $\varepsilon > 0$ such that for all $t' \geq T$ in this subsequence $M_{\sigma_{t'}}(b, C, S) \geq M^* + \varepsilon$ for all $(b, C, S) \in \overline{P}$. We know that there must exist a $t > T$ (not necessarily in the subsequence) such that $\bar{z}_t = (g, N^g, S)$ and, hence, $\overline{M}_{\sigma_t} = M^*$. Therefore, $M_{\sigma_t}(b, C, S) \leq M^*$ for all $(b, C, S) \in \overline{P}$. By Claim A.1, for all $(b, C, S) \in \overline{P}$,

$$\begin{aligned}
\frac{M_{\sigma_{t+1}}(b, C, S)}{M_{\sigma_t}(b, C, S)} &= \frac{\left(\frac{\frac{1}{t} + \sum_{C',S'|x_S(g,C')=x_S(b,C)} \sigma_t(g, C', S')}{\frac{1}{t} + \sum_{a',C',S'|x_S(a',C')=x_S(b,C)} \sigma_t(a', C', S')} \right)}{\left(\frac{\sum_{C',S'|x_S(g,C')=x_S(b,C)} \sigma_t(g, C', S')}{\sum_{a',C',S'|x_S(a',C')=x_S(b,C)} \sigma_t(a', C', S')} \right)} \\
&< \frac{\left(\frac{\frac{1}{t} + \sum_{C',S'|x_S(g,C')=x_S(b,C)} \sigma_t(g, C', S')}{\sum_{a',C',S'|x_S(a',C')=x_S(b,C)} \sigma_t(a', C', S')} \right)}{\left(\frac{\sum_{C',S'|x_S(g,C')=x_S(b,C)} \sigma_t(g, C', S')}{\sum_{a',C',S'|x_S(a',C')=x_S(b,C)} \sigma_t(a', C', S')} \right)} \\
&= \frac{\frac{1}{t}}{\sum_{C',S'|x_S(g,C')=x_S(b,C)} \sigma_t(g, C', S')} + 1,
\end{aligned}$$

which converges to 1 as $t \to \infty$ by Claim A.4. Therefore, for every $\delta > 0$, we can pick T large enough such that, for all $t \geq T$ such that $\bar{z}_t = (g, C, S)$,

$$\frac{M_{\sigma_{t+1}}(b, C, S)}{M_{\sigma_t}(b, C, S)} \leq 1 + \delta,$$

for all $(b, C, S) \in \bar{P}$. Finally, this means that we can also pick T and $t \geq T$ so that $\bar{z}_t = (g, C, S)$ and $M_{\sigma_{t+1}}(b, C, S) < M^* + \varepsilon$ for all $(b, C, S) \in \bar{P}$. Therefore, $M_{\sigma_{t+k}}(b, C, S) < M^* + \varepsilon$ for all $(b, C, S) \in \bar{P}$ and all $k \geq 1$, because by Claim A.2 the payoff of (b, C, S) is weakly decreasing when $M_{\sigma_t}(b, C, S) > M^*$. We, thus, reach a contradiction. □

Lemma A.3 and A.4 imply that $\lim_{t\to\infty} \overline{M}_{\sigma_t} = M^*$. Now, denote by Σ the set of limit points of σ_t.

LEMMA A.5. *All $\sigma \in \Sigma$ must induce the same joint distribution over (a, C), and this distribution must coincide with the unique equilibrium distribution.*

Proof. Note that $M_\sigma(z)$ is continuous in σ for all z. The previous conclusion implies that, for every $\sigma \in \Sigma$ and every z, $M_\sigma(z) \leq M^*$, with equality for $z \in Supp(\sigma)$. The equilibrium characterization results in Sections 3 and 4 established that every σ that satisfies this property induces the same distribution over (a, C). □

This completes the proof.

References

Ambuehl, Sandro and Heidi C. Thysen (2023). "Choosing between Causal Interpretations: An Experimental Study," Norwegian School of Economics,Working Paper.

Andre, Peter, Ingar Haaland, Christopher Roth and Johannes Wohlfart (2022). "Narratives about the Macroeconomy," Working Paper No. 18/21.

Ash, Elliot, Germain Gauthier and Philine Widmer (2024). "Text Semantics Capture Political and Economic Narratives," *Political Analysis*, 32(1), 115–132.

Ba, Cuimin (2023). "Robust Misspecified Models and Paradigm Shifts," University of Pennsylvania, Working Paper.

Bénabou, Roland, Armin Falk and Jean Tirole (2018). "Narratives, Imperatives, and Moral Reasoning.", NBER, Working Paper No. 24798.

Burstein, Paul (2003). "The Impact of Public Opinion on Public Policy: A Review and an Agenda," *Political Research Quarterly*, 56, 29–40.

Charles, Constantin and Chad W. Kendall (2023). "Causal Narratives," NBER, Working Paper No. 30346.

Cho, In-Koo and Kenneth Kasa (2015). "Learning and Model Validation," *Review of Economic Studies*, 82, 45–82.

Cowell, Robert G., Steffen L. Lauritzen, A. Philip Dawid, David J. Spiegelhalter, V. Nair, J. Lawless and M. Jordan (1999). *Probabilistic Networks and Expert Systems*, Berlin, Heidelberg: Springer-Verlag, 1st ed.

Eliaz, Kfir and Ran Spiegler (2020). "A Model of Competing Narratives," *American Economic Review*, 110(12), 3786–3816.

Esponda, Ignacio and Demian Pouzo (2017). "Conditional Retrospective Voting in Large Elections," *American Economic Journal: Microeconomics*, 9, 54–75.

Eyster, Erik and Michele Piccione (2013). "An Approach to Asset Pricing under Incomplete and Diverse Perceptions," *Econometrica*, 81, 1483–1506.

Healy, Andrew and Neil Malhotra (2013). "Retrospective Voting Reconsidered," *Annual Review of Political Science*, 16, 285–306.
Izzo, Federica, Gregory J. Martin and Steven Callander (2021). "Ideological Competition," SocArXiv. February, 19.
Jehiel, Philippe (2005). "Analogy-Based Expectation Equilibrium," *Journal of Economic Theory*, 123, 81–104.
Jones, Michael D., Mark K. McBeth and Elizabeth A. Shanahan (2014). "Introducing the Narrative Policy Framework," Palgrave Macmillan US, Chapter 1, 1–25.
Levy, Gilat and Ronny Razin (2021). "A Maximum Likelihood Approach to Combining Forecasts," *Theoretical Economics*, 16, 49–71.
Levy, Gilat, Ronny Razin and Alwyn Young (2022). "Misspecified Politics and the Recurrence of Populism," *American Economic Review*, 112(3), 928–62.
Lipset, Seymour M. and Stein Rokkan (1967). *Party systems and voter alignments: Cross-national perspectives*, vol. 7, New York: Free Press.
Macaulay, Alistair (2022). "Shock Transmission and the Sources of Heterogeneous Expectations," University of Oxford, Working Paper No.
Mailath, George J. and Larry Samuelson (2020). "Learning under Diverse World Views: Model-Based Inference," *American Economic Review*, 110(5), 1464–1501.
Montiel, Olea, Jose Luis, Pietro Ortoleva, Mallesh M. Pai and Andrea Prat (2022). "Competing Models," *Quarterly Journal of Economics*, 137, 2419–2457.
Pearl, Judea (2009). *Causality: Models, Reasoning and Inference*, USA: Cambridge University Press, 2nd ed.
Persson, Torsten and Guido Tabellini (2000). *Political Economics: Explaining Economic Policy*, MIT press.
Plescia, Carolina and Sylvia Kritzinger (2017). "Retrospective Voting and Party Support at Elections: Credit and Blame for Government and Opposition," *Journal of Elections, Public Opinion and Parties*, 27, 156–171.
Polletta, Francesca (2008). "Storytelling in politics," *Contexts*, 7, 26–31.
Roemer, John E. (1994). "The Strategic Role of Party Ideology when Voters are Uncertain about How the Economy Works," *American Political Science Review*, 88, 327–335.
Rothschild, Michael and Joseph Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90, 629–649.
Sanders, Karen, María Jesús Molina Hurtado and Jessica Zoragastua (2017). "Populism and Exclusionary Narratives: The 'Other' in Podemos' 2014 European Union Election Campaign," *European Journal of Communication*, 32, 552–567.
Schnellenbach, Jan and Christian Schubert (2015). "Behavioral Political Economy: A Survey," *European Journal of Political Economy*, 40, 395–417.
Schwartzstein, Joshua and Adi Sunderam (2021a). "Shared Models in Networks, Organizations, and Groups," Working Paper No. 30642, Harvard University.
Schwartzstein, Joshua and Adi Sunderam (2021b). "Using Models to Persuade," *American Economic Review*, 111(1), 276–323.
Shanahan, Elizabeth A., Mark K. McBeth and Paul L. Hathaway (2011). "Narrative Policy Framework: The Influence of Media Policy Narratives on Public Opinion," *Politics & Policy*, 39, 373–400.
Shiller, Robert J. (2017). "Narrative Economics," *American Economic Review*, 107(4), 967–1004.
Spiegler, Ran (2013). "Placebo Reforms," *American Economic Review*, 103(4), 1490–1506.
Spiegler, Ran (2016). "Bayesian Networks and Boundedly Rational Expectations," *Quarterly Journal of Economics*, 131, 1243–1290.
Spiegler, Ran (2020). "Behavioral Implications of Causal Misperceptions," *Annual Review of Economics*, 12, 81–106.

Stone, Deborah A. (1989). "Casual Stories and the Formulation of Agendas," *Political Science Quarterly*, 104, 282.
Szeidl, Adam and Ferenc Szűcs (2022). *The Political Economy of Alternative Realities*, Centre for Economic Policy Research.
Weaver, R. Kent (2013). "Policy Leadership and the Blame Trap: Seven Strategies for Avoiding Policy Stalemate," *Governance Studies*, Brookings Institution.

Journal of the European Economic Association 2025 23(3):983–1027
https://doi.org/10.1093/jeea/jvae047